

**Competing risks analysis of end-stage-renal disease and mortality  
among adults with diabetes - a comparison of First Nations people and  
other Saskatchewan residents**

**Submitted to  
The Department of Community Health & Epidemiology  
In partial fulfillment  
of the requirements for the degree of  
Master of Science**

**By  
Ying Jiang**

**© Copyright Ying Jiang, April 2012. All rights reserve**

## **PERMISSION TO USE**

In presenting this thesis / dissertation in partial fulfillment of the requirements for a postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/ dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or profession who supervised my thesis/ dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work has done. It is understood that any copying or publication or use of this thesis/ dissertation or parts there of for financial gain shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this  
thesis/dissertation in whole or part should be addressed to:

Head of the Department of Community Health and Epidemiology  
University of Saskatchewan  
Saskatoon, Saskatchewan S7N 0W8

## **Abstract**

**Background:** End stage renal disease (ESRD) is a growing public health problem in Canada and it disproportionately affects Aboriginal people. Diabetes is the most common reported cause of ESRD.

**Objectives and methods:** To determine whether there are significant disparities in the risk of ESRD and mortality without ESRD between diabetic First Nations (FN) and other Saskatchewan (OSK) people; to build and validate diabetic ESRD dynamic models. This is a population study of diabetes, utilizing data drawn from the Saskatchewan Ministry of Health administrative databases from 1980 to 2005. Competing risks survival analysis was used, including a Cox cause-specific model, Weibull proportional hazards (PH) model and piece-wise exponential PH hazards model. System Dynamics modeling (SDM) and agent-based modeling (ABM) methods were used to build dynamic models of diabetic patients' progression to ESRD.

**Results:** There were a total of 90,429 diabetic people in the study cohort, from 1980 to 2005. Among them, 8,254 (9%) of them were FN people. The average age at diabetes diagnosis for FN was 47.2 (SD=14) years old while for OSK, it was 61.6 (SD=15.3) years old (P-value<0.0001). After adjusting for sex and age at diabetes diagnosis, the risk of developing ESRD was 2.97 times higher for FN compared to OSK (95% CI: 2.51-3.54; P-value<0.0001). FN had lower risk of death than OSK before adjusting for age and sex difference. After adjusting for diabetes diagnosis age, sex, interaction between age and sex and interaction between age and ethnicity, FN had higher risk of death than OSK given the same sex and diabetes diagnosis age (younger than 81 years old). Using the same hazard rate estimations from competing risks survival analysis, the ABM model demonstrated a better match between historical data and model predicted data compared to the SD model.

**Conclusion:** A much younger age of diabetes diagnosis among FN compared to OSK likely contributes to higher rates of ESRD because of a differential mortality effect – FN with diabetes are more likely to live long enough to develop ESRD.

## Table of contents

Chapter 1 Introduction .....	1
Chapter 2 Literature .....	3
2.1 Nomenclature .....	3
2.2 Epidemiology of diabetes .....	4
2.2.1 Introduction .....	4
2.2.2 General background.....	5
2.2.3 Epidemiology of diabetes among Aboriginal people in North America .....	6
2.2.4 Etiology and risk factors for Type 1 diabetes .....	7
2.2.5 Etiology and risk factors for Type 2 diabetes .....	8
2.2.6 Etiology and risk factors for gestational diabetes.....	10
2.3 Epidemiology of CKD .....	11
2.3.1 Introduction .....	11
2.3.2 Epidemiology of CKD.....	12
2.3.3 Complications of CKD .....	13
2.3.4 Epidemiology of CKD among Aboriginal population. ....	13
2.3.5 Risk factors for CKD.....	15
2.3.6 Summary.....	17
2.4 System Dynamics modeling and agent-based modeling for public health .....	17
Chapter 3 Rationale.....	23
Chapter 4 Study objectives .....	25
Chapter 5 Methodology .....	26
5.1 Study design and participants.....	26
5.2 Statistical analysis .....	27
5.2.1 Introduction .....	27
5.2.2 Non-parametric methods .....	30
5.2.3 Cox proportional hazards model.....	32
5.2.4 Parametric proportional hazards model.....	33
5.2.5 Competing risks survival analysis .....	35

5.3 System Dynamics modeling.....	38
5.3.1 System Dynamics modeling, first order delay and competing risks .....	38
5.3.2 Subscribing Dynamics Modeling .....	43
5.4 Agent-based modeling.....	44
5.5 Software .....	46
Chapter 6 Results .....	47
6.1 Description of study population and descriptive analysis.....	47
6.2 Description of variables .....	53
6.3 Non-parametric model.....	53
6.3.1 Cause-specific hazard rates .....	53
6.3.2 Cumulative incidence function curves .....	55
6.4 Cox cause-specific model.....	56
6.5 Parametric proportional hazards model.....	59
6.6 Piece-wise exponential proportional hazards model.....	62
6.7 Computational Modeling .....	66
6.7.1 System Dynamics modeling .....	66
6.7.2 Agent-based modeling.....	72
Chapter 7 Discussion and conclusions.....	86
7.1 Interpretation .....	86
7.2 Limitations and strengths .....	89
7.2.1 Limitations of the study population.....	89
7.2.2 Strengths of the study population. ....	91
7.3 Discussion of methodology.....	92
7.4 Conclusions .....	95
7.5 Future work .....	95

## List of Figures

Figure 2.1 Age-standardized diabetes incidence by ethnicity and sex (23).....	7
Figure 2.2 Age-standardized diabetes prevalence by ethnicity and sex (23).....	7
Figure 2.3 Age-standardized ESRD incidence among subjects with diabetes (6).....	14
Figure 2.4 Age-standardized ESRD prevalence among subjects with diabetes (6).....	14
Figure 2.5 Overview of model structure with population stocks and flows and policy related intervention (shown in <i>italics</i> ) (13).....	19
Figure 2.6 Structure of agent-based model of TB diffusion with the effect of smoking (66).....	21
Figure 5.1 Competing risk events. ....	36
Figure 5.2 Stock and flow .....	38
Figure 5.3 First order delay for progression from diabetes to ESRD .....	39
Figure 5.4 Competing risks between ESRD and death without ESRD. ....	42
Figure 5.5 Subgroup in Vensim .....	44
Figure 5.6 Statechart. ....	45
Figure 6.1 Subjects in study. ....	49
Figure 6.2 ESRD cases by ethnicity.....	50
Figure 6.3 Death cases by ethnicity. ....	51
Figure 6.4 Hazard rates for ESRD and death cases. ....	54
Figure 6.5 Hazard rates of ESRD by ethnicity.....	55
Figure 6.6 Hazard rates of death without ESRD by ethnicity.....	55
Figure 6.7 Cumulative incidence function (CIF) curve for ESRD by ethnicity. ....	56
Figure 6.8 Cumulative incidence function (CIF) curve for death by ethnicity.....	56
Figure 6.9 System Dynamics Model with a 5-year interval. ....	68
Figure 6.10 Comparison of historical data with model-estimated data with a 3-year interval analysis of a piece-wise exponential model for ESRD.....	69
Figure 6.11 Comparison of historical data with model-estimated data with a 5-year interval analysis of a piece-wise exponential model for ESRD.....	70
Figure 6.12 Comparison historical data with model estimated data with 3-year interval analysis of piece wise exponential for death without ESRD. ....	70
Figure 6.13 Comparison of historical data with model-estimated data with a 5-year interval analysis of a piece-wise exponential model for death without ESRD.....	71
Figure 6.14 State chart of the agent-based model for person flow. ....	72
Figure 6.15 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for ESRD. ....	74
Figure 6.16 Comparison between historical data and model-estimated data with a 5-year interval analysis of a piece-wise exponential model for ESRD. ....	74
Figure 6.17 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for death without ESRD. ....	75
Figure 6.18 Comparison between historical data and model-estimated data with a 5-year interval analysis of a piece-wise exponential model for death without	

ESRD. ....	75
Figure 6.19 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for ESRD in FN males. .....	76
Figure 6.20 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for ESRD in FN females. ....	77
Figure 6.21 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for ESRD in OSK males. ....	77
Figure 6.22 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for ESRD in OSK females. ....	78
Figure 6.23 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for death without ESRD in FN males. ....	78
Figure 6.24 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for death without ESRD in FN females. ....	79
Figure 6.25 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for death without ESRD in OSK males. ....	79
Figure 6.26 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential for death without ESRD in OSK females. ....	80
Figure 6.27 Comparison between incident historical data and model-estimated data for ESRD from a Monte Carlo model. ....	81
Figure 6.28 Comparison between incident historical data and model-estimated data for death without ESRD from a Monte Carlo model. ....	81
Figure 6.29 Comparison between incident historical data and model-estimated data for ESRD in FN males from a Monte Carlo model. ....	82
Figure 6.30 Comparison between incident historical data and model-estimated data for ESRD in FN females from a Monte Carlo model. ....	82
Figure 6.31 Comparison between incident historical data and model-estimated data for ESRD in OSK males from a Monte Carlo model. ....	83
Figure 6.32 Comparison between incident historical data and model-estimated data for ESRD in OSK females from a Monte Carlo model. ....	83
Figure 6.33 Comparison between incident historical data and model-estimated data for death without ESRD in FN males from a Monte Carlo model. ....	84
Figure 6.34 Comparison between incident historical data and model-estimated data for death without ESRD in FN females from a Monte Carlo model. ....	84
Figure 6.35 Comparison between incident historical data and model-estimated data for death without ESRD in OSK males from a Monte Carlo model. ....	85

Figure 6.36 Comparison between incident historical data and model-estimated data for death without ESRD in OSK females from a Monte Carlo model. ....	85
Figure 7.1 Confounding factor.....	90

## List of Tables

Table 6.1. Baseline characteristics by ethnicity. ....	52
Table 6.2. Description of variables. ....	53
Table 6.3. Univariable Cox cause-specific model for ESRD outcome.....	57
Table 6.4. Univariable Cox cause-specific model for Death outcome.....	57
Table 6.5. Multivariable Cox cause-specific model for ESRD outcome. ....	58
Table 6.6. Multivariable Cox cause-specific model for death outcome.....	58
Table 6.7. AIC in Parametric PH models for ESRD. ....	59
Table 6.8. AIC in Parametric PH for death without ESRD.....	59
Table 6.9. Weibull PH model for ESRD. ....	61
Table 6.10. Weibull PH model for death without ESRD. ....	61
Table 6.11. Piece-wise exponential model for ESRD outcome with 3-year interval...	64
Table 6.12. Piece-wise exponential model for ESRD outcome with 5-year interval. .	64
Table 6.13. Piece-wise exponential model for death outcome with 3-year interval. ...	65
Table 6.14. Piece-wise exponential model for death outcome with 5-year interval. ...	66



## **Abbreviations**

<b>BMI</b>	<b>Body Mass Index</b>
<b>CKD</b>	<b>Chronic Kidney Disease</b>
<b>CG</b>	<b>Cockcroft-Gault</b>
<b>CCHS</b>	<b>Canadian Community Health Survey</b>
<b>CIHI</b>	<b>Canadian Institute for Health Information</b>
<b>ESRD</b>	<b>End Stage Renal Disease</b>
<b>eGFR</b>	<b>Estimated Glomerular Filtration Rate</b>
<b>FN</b>	<b>First Nations</b>
<b>GDM</b>	<b>Gestational Diabetes Mellitus</b>
<b>IGT</b>	<b>Impaired Glucose Tolerance</b>
<b>IFG</b>	<b>Impaired Fasting Glucose</b>
<b>MDRD</b>	<b>Modification of Diet in Renal Disease Study</b>
<b>NHANES</b>	<b>National Health and Nutrition Examination Survey</b>
<b>OSK</b>	<b>Other Saskatchewan People</b>
<b>RHS</b>	<b>First Nations Regional Longitudinal Health Survey</b>
<b>T1DM</b>	<b>Type 1 Diabetes Mellitus</b>
<b>T2DM</b>	<b>Type 2 Diabetes Mellitus</b>
<b>WHO</b>	<b>World Health Organization</b>

## **Chapter 1 Introduction**

End stage renal disease (ESRD) occurs when the kidneys can no longer function adequately for an individual's daily life. ESRD is considered the last stage of chronic kidney disease (CKD), in which only approximately 10% of kidney function remains (1). Patients with ESRD need medical treatment to survive, such as dialysis or a kidney transplant. ESRD is a growing public health problem in Canada. It was estimated that almost 38,000 people were being treated for ESRD by the end of 2009 (2). Diabetes is the most common cause of ESRD, accounting for 53.8% of incident cases in 2009, up from 26.9% in 1996 (2, 3).

ESRD disproportionately affects Aboriginal people in Canada (4-6). For example, Young et al. (4) found that between 1981 and 1986, the age-standardized incidence of ESRD among Canadian Aboriginal people was more than 2.5 times higher than that found in the general population. Moreover, Dyck et al. (6) found that from 1980 to 2005 in Saskatchewan, First Nations (FN) people had higher rates of diabetes-related ESRD than other Saskatchewan (OSK) people. The reasons for the greater incidence of diabetic ESRD among Aboriginal people with diabetes compared to other Canadians with diabetes is currently unclear, but it may result from increased onset rates of diabetic glomerulosclerosis, which may result in ESRD (7), higher rates of kidney function loss after onset of CKD (CKD) (7), increased survival during later stages of CKD (7), poorer glycemic control and reduced access to treatment that helps to slow

the progression of CKD (7, 8). In this study, we examined whether there are differences in the risks of developing ESRD and death without ESRD between First Nations (FN) and other Saskatchewan people (OSK) using competing risks survival analysis. Estimations of simultaneous risks of ESRD and death without ESRD will help to understand ethnic differences in the rates of ESRD. A better understanding of ethnicity-related differences in the risks of these two events would potentially help to inform research and interventions that are aimed at reducing disparities in the ESRD incidence between First Nations people and other Canadians, and it could be used to predict future renal replacement requirements and costs (9).

In recent years, public health is starting to use computational modeling and simulation experiments to assist stakeholders and policy makers to better understand the health care system's structure and interactions among components, to project possible future burden of disease outcomes and expenditures, to evaluate health care goals and planning strategies both in the short term and the long term, and to understand whether there are any conflicts among goals and then to make robust policy (10-15). In this thesis, both System Dynamics modeling (SDM) and agent-based modeling (ABM) will be used to build diabetic ESRD models and validate these models by comparing historical data with model-predicted data.

The thesis will be divided into the following sections: literature review, including the epidemiology of diabetes, epidemiology of CKD, and the use of SDM and ABM for public health research; rationale; study objectives; methodology; results; discussion; and conclusions.

## **Chapter 2 Literature**

### **2.1 Nomenclature**

We will use similar terminology described by Young about indigenous populations in this thesis (16): “The term Native American encompasses North American Indians, Eskimos (Inuit) and Aleut ... In Canada, the term Native continues to be used by some Native Organizations, although Aboriginal seems to be preferred. Three aboriginal groups are recognized in Canada: Indians, Inuit and Métis. The term Indian is slowly being replaced by First Nation ... The term Métis is used only in Canada, and refers to a distinct cultural group that originated from mixed Indian-White marriages in the early settlement of the Canadian West.”

Throughout this thesis, Aboriginal will be used to refer to all North American Indigenous People including people of Indian, Inuit and Métis heritage, as well as non-Indigenous registered Indians. First Nation will be used to refer to the registered indigenous population in Canada.

## **2.2 Epidemiology of diabetes**

### **2.2.1 Introduction**

Diabetes mellitus (DM) is a chronic disease caused by insufficiencies in insulin, compromised action of insulin or some combination of the two. The World Health Organization (17) classifies DM into three types: type 1 diabetes (T1DM), type 2 diabetes (T2DM) and gestational diabetes mellitus (GDM). Diabetes disproportionately affects some ethnic groups, such as Aboriginal, African-American, and Hispanic people (18).

Diabetes causes many chronic complications, such as cardiovascular disease, eye disease, and renal failure. In 2005, almost 1.1 million people died from diabetes, with 80% of deaths occurring in low- and middle-income countries (18). According to the International Diabetes Federation (19), in 2010, almost 4 million deaths across the world could be attributed to diabetes, accounting for 6.8% of all deaths among people aged 20-79 years. Globally, by 2030, the number of people with diabetes is expected to double (19). Because diabetes' complications are long term, treatment for diabetes and its complications are costly. Global healthcare expenditures to treat and prevent diabetes and its complications were estimated to be at least 376 billion US dollars (USD) in 2010 and projected to be USD 490 billion by 2030 (20).

### **2.2.2 General background**

In Canada in 2006-07, approximately 2 million Canadians, or 6.2% of the Canadian population, were diagnosed with diabetes – an increase of 21% from 2002-03 (21). There were 211,168 incident cases of diabetes reported in Canada between 2006 and 2007 (21). The age-standardized prevalence of diagnosed diabetes varies geographically in Canada, with higher prevalence in several of the eastern provinces (e.g., Newfoundland and Labrador, Nova Scotia, and New Brunswick) as well as Manitoba and lower prevalence in the west (e.g., Alberta, British Columbia, and Saskatchewan) (21).

Impaired glucose tolerance (IGT) and impaired fasting glucose (IFG) are considered to be intermediate conditions between normal blood glucose levels and diabetes. They will also be referred to as the prediabetes stage in later parts of this thesis. T1DM is insulin dependent and results from a lack of insulin. Although T1DM occurs mostly among children and young adults, it may appear among all age groups (18). T2DM results from insulin resistance and eventually from a decrease in insulin efficiency. Factors that are related to T2DM may include age, obesity, physical inactivity, diet, insulin resistance, family history of diabetes and GDM (19). A diagnosis of T2DM is more common after 40 years of age, but some cases are diagnosed earlier.

### **2.2.3 Epidemiology of diabetes among Aboriginal people in North America**

Aboriginal people are more likely to develop T2DM than other Canadians. The First Nations Regional Longitudinal Health Survey (RHS) (2002/2003) reported the prevalence of diagnosed diabetes among FN adults to be 19.7%, with most of the cases being T2DM (22). FN people experience a higher prevalence of diabetes than the general population in every age group. In addition, the RHS report described diabetic people on reserves to be much younger than other Canadians with diabetes; that is, 65% of FN diabetic people on reserves were 45 years old or younger, whereas the average age of diabetic people in the general population is typically over 60 years of age.

Dyck et al. (23) studied the epidemiology of diabetes among FN and OSK between 1980 and 2005 in Saskatchewan and reported a higher incidence and prevalence of diabetes among FN people, shown in Figure 2.1 and

Figure 2.2. It has been shown that most new cases of diabetes in FN people occur at the age of 40-49 years; however, most new cases of diabetes in OSK people appear among those 70 or older. Similar findings have been reported in other Canadian provinces. For example, in Alberta from 1995 to 2005 (24), the age- and sex-adjusted prevalence of diabetes among FN people was approximately twice that of non-FN people. Similarly, in Manitoba between 1989 and 1998 (25), the age-standardized prevalence of diabetes was 4.5 times higher among FN people than among non-FN people.

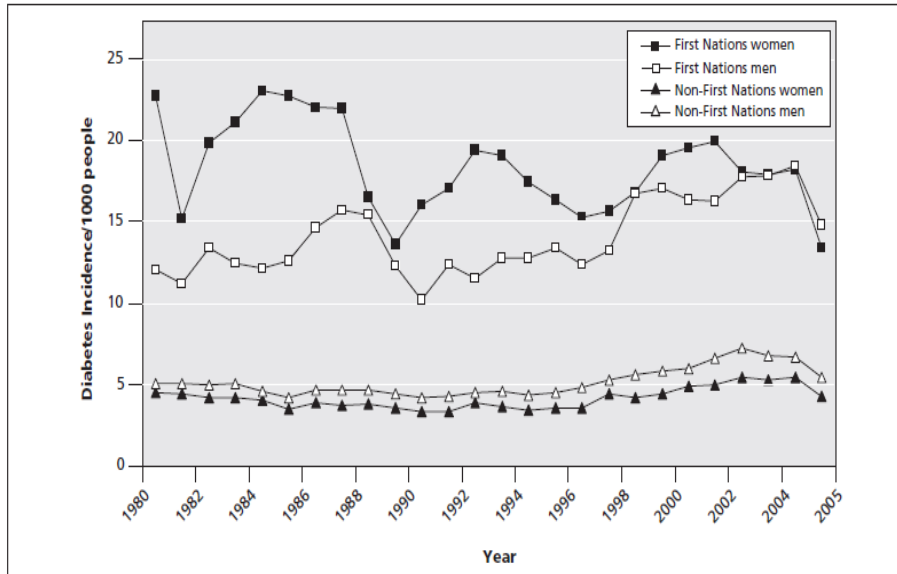


Figure 2.1 Age-standardized diabetes incidence by ethnicity and sex (23).

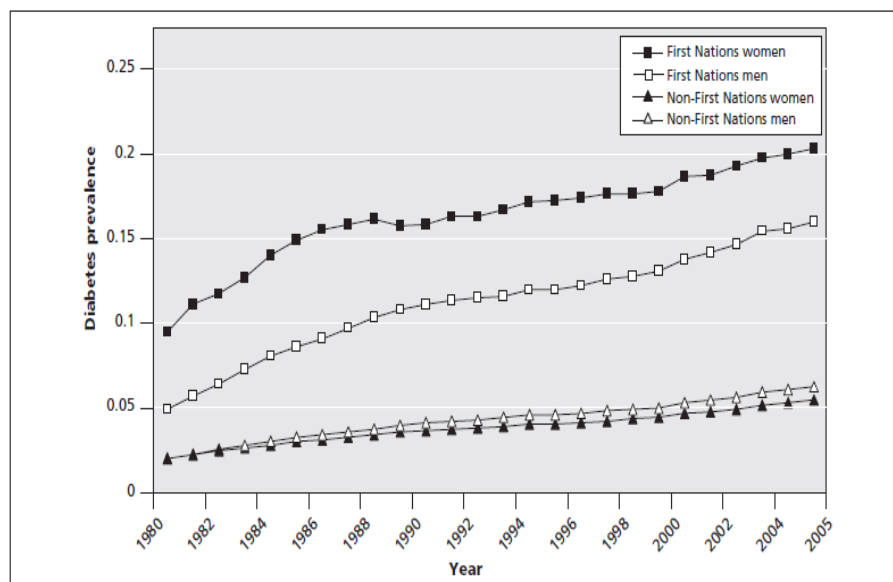


Figure 2.2 Age-standardized diabetes prevalence by ethnicity and sex (23).

#### 2.2.4 Etiology and risk factors for Type 1 diabetes

T1DM and T2DM involve different underlying mechanisms. T1DM is “characterized by reduced beta cell numbers in the pancreas, which results in insulin production deficiency and is often caused by T-cell’s autoimmune attack, which leads



to beta cell loss” (26). Genetic factors associated with the development of T1DM may include human leukocyte antigen class II genes (26). A variety of environmental factors have been associated with the development of T1DM, such as viruses, insufficient breastfeeding duration and increased maternal age. Coxsackie virus B infection is proposed to have effects on the pancreas and may also be related to T1DM (26). Early diet may play a role in development of T1DM later in life. For example, Harrison and Honeyman (27) found that children who were breastfed for a longer period had a lower risk of developing T1DM.

### **2.2.5 Etiology and risk factors for Type 2 diabetes**

For T2DM, risk factors may include age, overweight, obesity, physical inactivity, IGT, GDM, hypertension, and decreased high-density lipoprotein cholesterol (28). The importance of genetic and environmental factors in the etiology of diabetes varies among populations (29). It was suggested recently that the dramatic increases in the rates of T2DM can be primarily attributed to increases in obesity (30). In addition to general obesity, body fat distribution is considered another risk factor for T2DM, particularly “central obesity”. Central obesity is also known as abdominal obesity, characterized by an accumulation of fat around the waist (31). Recently, the relationship between excess BMI duration and diabetes incidence was examined by Lee et al.(32). One longitudinal study was performed with 8,157 adolescents and young adults aged 14 to 21 years from 1981 to 2006. Self-reported diabetes status, height and weight were recorded for each participant. Excess BMI-years was the

exposure factor. To calculate this indicator, actual BMI was first subtracted from the reference BMI for each person in each study year, and then the excess BMI for the whole study period was calculated. The results from that study showed that long excess BMI duration increases the risk of diabetes.

A high prevalence of obesity has been documented among many Aboriginal populations. The 2004 Canadian Community Health Survey (CCHS) (33) indicated that 23.1% of Canadians aged 18 years and older were obese, compared to 13.8% in 1978/1979. In the same study, the proportion of Aboriginal people who were obese was 37.6%, almost 1.6 times higher than that among the general population.

Regular physical activity is considered helpful in preventing and managing T2DM. Helmrach et al. (34) examined physical activity and its relationship with later development of T2DM. They reported that incidence rates of diabetes decreased as physical activity increased. The protective effect of physical activity was strongest among people who were at greatest risk of T2DM, such as those with a parental history of T2DM, a high body mass index (BMI) and a history of hypertension.

Investigations with Aboriginal people have also shown that a traditional lifestyle may be protective against the development of T2DM. For example, Schulz et al. (35) assessed the relationships between T2DM and risk factors such as obesity, physical activity, and diet among Pima Indians in Mexico compared to Pima Indians in Arizona. These two Pima groups are closely related to each other and share extensive genetic similarity. The results indicated that compared to U.S. Pima Indians, Mexican Pima Indians have significantly higher physical activity levels and a diet with a lower

percentage of calories. The age- and sex-adjusted prevalence of T2DM for Mexican Pima Indians was less than 20% of that in U.S. Pima Indians, which provides evidence that loss of a traditional lifestyle and adoption of a Western lifestyle may be associated with the T2DM epidemic.

#### **2.2.6 Etiology and risk factors for gestational diabetes**

GDM is defined as “carbohydrate intolerance with first onset or recognition during pregnancy” (26). In 2002, Dyck et al. (36) compared rates of GDM between Aboriginal and non-Aboriginal women in the Saskatoon Health District, reporting the prevalence of GDM as 3.5% and 11.5% for women from the non-Aboriginal population and Aboriginal population, respectively. Aboriginal ethnicity was an independent predictor for the development of GDM.

Common risk factors identified for GDM included being pregravid overweight or obese, previous pregnancy with GDM, IGT, advanced maternal age, a family history of T2DM, and previous delivery of children with a high birth weight (36, 37). GDM may result in maternal and fetal risks, such as recurrent GDM (38), early development of diabetes for both mother and offspring (39), and macrosomia (40). Pettitt et al. (41) investigated the long-term effects of GDM on the offspring of Pima Indians. The findings from this study suggest that intrauterine exposure to diabetes may be an important factor for the development of diabetes later in life. Osgood et al. (42) investigated the effect of GDM on the development of T2DM in Saskatchewan by constructing a population-based simulation model to understand the inter/intra-

generational interactions of GDM and T2DM. This study suggested that GDM contributed to between 19% and 30% of T2DM among FN in Saskatchewan.

## **2.3 Epidemiology of CKD**

### **2.3.1 Introduction**

The kidneys help to excrete waste products arising from normal body functions. These waste products include creatinine which is produced from muscle metabolism. Because serum creatinine is completely filtered and excreted by normal kidneys, the serum creatinine level can be used to estimate the glomerular filtration rate (GFR), which is defined as the per-minute filtered output produced by the kidneys' glomeruli. Thus, the GFR will decrease when kidney function is impaired and this will be detectable by an increase in serum creatinine (43). Using serum creatinine to estimate GFR, chronic kidney disease can be divided into five stages (44):

“Stage 1: kidney damage is indicated by leakage of small amounts of albumin into the urine (microalbuminuria), but there is a normal or above-normal glomerular filtration rate ( $\text{GFR} \geq 90 \text{ mL/min/1.73 m}^2$ ).

Stage 2: kidney damage is often indicated by increasing microalbuminuria. GFR is slightly decreased at ranges from 60-89  $\text{mL/min/1.73 m}^2$ .

Stage 3: Microalbuminuria may have progressed to macro-albuminuria. There is a moderate reduction in GFR (30-59  $\text{mL/min/1.73 m}^2$ ).

Stage 4: kidney damage worsens and GFR ranges from 15-29  $\text{mL/min/1.73 m}^2$ .

Stage 5: severe kidney failure ( $\text{GFR} < 15 \text{ mL/min/1.73 m}^2$ ). Treatments such as

hemodialysis, peritoneal dialysis and kidney transplantation are needed for patients to survive.”

To monitor the progression of CKD, routine measurement of serum creatinine is needed. By early detection and treatment, patients with CKD may delay the progression of CKD and prevent adverse outcomes related to CKD (8). The prevalence of CKD can be determined by using equations that estimate GFR. The Modification of Diet in Renal Disease Study (MDRD) equation and the adjusted Cockcroft-Gault (CG) equation have been used in different studies, which increases the difficulty of evaluating and comparing epidemiological articles. Zhang et al. (45) systematically reviewed population-based studies of the prevalence of CKD and concluded that the MDRD equation is used more frequently than the CG equation in recent epidemiology studies and stated that the precision of the equations in estimating GFR needs to be further investigated.

### **2.3.2 Epidemiology of CKD**

In 2008, an estimated 36,638 Canadians were receiving treatment for ESRD. Recent Canadian costing studies suggest that the annual cost per ESRD patient is 55,466 Canadian dollars (46). Diabetes is the most common cause of ESRD, accounting for 53.8% of incident cases in 2009, up from 26.9% in 1996 (2, 3). In the United States, Coresh et al. (47) compared two samples from the National Health and Nutrition Examination Survey (NHANES) (NHANES 1988-1994 and NHANES 1999-2004) and found that the prevalence of CKD had increased; that is, the prevalence of

individuals with stage 1-4 CKD increased from 10% during 1988-1994 to 13.1% during 1999-2004. It was proposed that the increase in CKD may be due to concomitant increases in the prevalence of diabetes and hypertension.

### **2.3.3 Complications of CKD**

Complications that may develop during CKD include hypertension, malnutrition, bone disease, and anemia (44). Mortality rates from coronary artery disease and strokes are high among patients with late stages of CKD (stages 3 and 4) , and most patients die without developing ESRD (48, 49). Dialysis and renal transplantation are initially lifesaving after the development of ESRD. However, estimates from the US Renal Data System showed that for patients who started dialysis in 2001, the 3-year survival probability was only 54% (50). In Saskatchewan, it was estimated that the median survival time after the diagnosis of diabetic ESRD was only 2 years (6). Among patients with pre-ESRD CKD, death is a more common complication than is ESRD (51).

### **2.3.4 Epidemiology of CKD among Aboriginal population.**

Aboriginal people experience higher rates of ESRD than non-Aboriginal people. In Saskatchewan in 1994, Dyck and Tan (52) found that the incidence of diabetic ESRD was higher among Aboriginal people than non-Aboriginal people during a 10-year period. Similarly, Nelson et al (53) found that among 45-64-year-olds, the incidence of ESRD among diabetic Pima Indians was 14 times higher compared with the rate among

the United States diabetic population. One recent study in Saskatchewan showed that FN with diabetes had higher rates of ESRD when compared with OSK with diabetes (Figure 2.3 and Figure 2.4) and that FN were significantly younger when diagnosed than were OSK (6). Among OSK people, males experienced higher rates of ESRD than did females, while the difference in ESRD incidence between FN males and FN females varied during the study period.

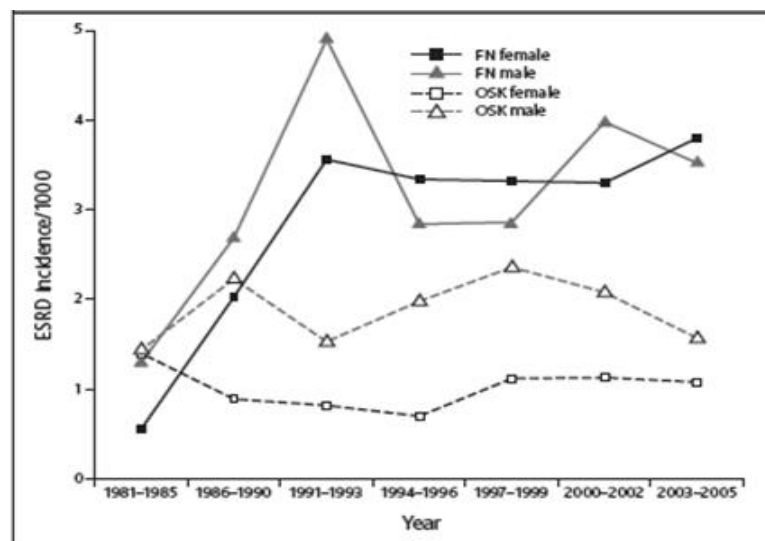


Figure 2.3 Age-standardized ESRD incidence among subjects with diabetes (6).

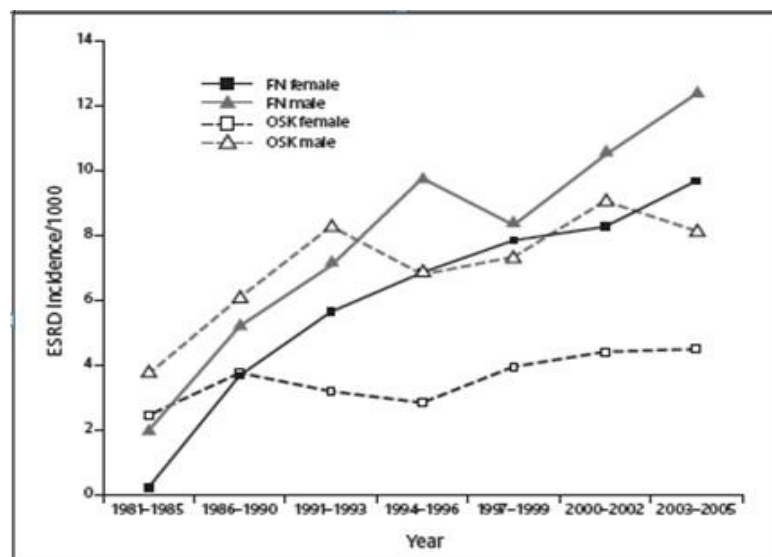


Figure 2.4 Age-standardized ESRD prevalence among subjects with diabetes (6).

The difference between Aboriginal populations and non-Aboriginal populations in the prevalence of CKD varies among stages of CKD (8). Gao et al. (8) have shown the overall age- and sex-adjusted prevalence of CKD to be higher among FN compared with non-First Nations Canadians. After stratifying CKD by stages, they found that age- and sex-adjusted rates of CKD are even higher among First Nations for the more severe stages of CKD.

Ethnic differences in the prevalence of CKD between African-Americans and non-African-Americans were examined in one cohort study in the United States (54). The MDRD equation for estimated GFR (eGFR) was used, and this study showed that after adjustment for other individual factors, the black-to-white odds ratio was 0.42 at an eGFR of 50 to 59 ml/min/ 1.73 m<sup>2</sup> and increased to 1.73 at an eGFR of 10 to 19 ml/min per 1.73 m<sup>2</sup>.

### **2.3.5 Risk factors for CKD**

Both genetic and environmental factors may play roles in the development of CKD (55, 56), and they may include age, hypertension, diabetes, obesity and smoking. One 10-year cohort study performed by Yamagata et al (57) has shown several risk factors for the development of CKD. Among 123,764 people aged 40 years and older, for females, the development of CKD stage 1 or 2 was most strongly related to age, hypertension, IGT, diabetes, obesity, and current smoking. For males, the development of CKD stage 1 or 2 was most strongly associated with the same factors in addition to



alcohol consumption.

Regarding genetic factors, there is evidence for family clustering of diabetic renal disease among Aboriginal populations (58). Pettitt et al (58) found that after adjusting for sex and other risk factors, proteinuria occurred among 22.9% of diabetic offspring with at least one parent having diabetes and 45.9% of diabetic offspring with both parents having diabetes. There is also a relationship between high blood pressure and diabetic nephropathy (44). Genes that encode proteins that are related to hypertension might be associated with diabetic nephropathy (44). It was proposed that the angiotensin I-converting enzyme gene is an independent risk factor for renal disease and is associated with an increased rate of diabetic kidney disease progression (59, 60).

Regarding environmental factors, intrauterine exposure to diabetes may be relevant to diabetic renal disease among mothers and their offspring (61). Mothers who have GDM may develop T2DM and diabetic ESRD later in life. The offspring of mothers with GDM may have a higher risk of diabetes when they reach childbearing ages, and they have a higher risk of developing diabetic renal disease than those not exposed to intrauterine diabetes. Nelson et al. (61) found that the risk of increased urinary albumin excretion among offspring of diabetic mothers was nearly four times that among offspring of non-diabetic or pre-diabetic mothers. Moreover, Dyck et al. (62) reported that diabetic pregnancy may increase the risk of diabetic ESRD and that intrauterine exposure to diabetes may be one environmental factor for the development of diabetic renal disease among mothers and their offspring.

### **2.3.6 Summary**

Aboriginal populations experience higher rates of CKD compared with non-Aboriginal populations. Risk factors for CKD may include age, hypertension, IGT, diabetes, obesity and smoking. Diabetes is reported to be the most common cause of ESRD. Patients with ESRD need to take costly treatments such as dialysis and kidney transplantation. Death rates are reported to be high among CKD patients.

## **2.4 System Dynamics modeling and agent-based modeling for public health**

System Dynamics modeling (SDM) is suitable to help understand complex dynamic behaviors of health care systems and health of the population (11). SDM simulates the systems at an aggregate level with the inclusion of characteristics such as causal loops, delays and interactions among components of the modeled system (15). Agent-based modeling (ABM) is useful to help in understanding the behaviors of autonomous agents and the emergent performance of the system in which these individuals circulate (63). Once either System Dynamic models or agent-based models have been validated, we can perform simulation experiments to ask certain “what if” questions (10, 15). For example, with diabetic ESRD models, it is possible to understand how the number of ESRD cases might change if we were to lower diabetic mortality.

SDM has been used in public health to understand the evolution of complex health systems and health of populations over time (11). With a validated System

Dynamic model, experiments under certain scenarios could be conducted. For example, Jones et al. (13) used system dynamic modeling to understand diabetes population dynamics. The simulation results help understand the effects of factors such as obesity and clinical management of prediabetes and diabetes on the change of the prevalence of diabetes. Figure 2.5 presents an overview of the System Dynamics model of the diabetic population in Jones et al. (13). The whole population is divided into several stages according to blood glucose level and diagnosis status. Flows are represented as arrows such as the progression from undiagnosed, uncomplicated diabetes to undiagnosed complicated diabetes. This model also includes factors related to interventions. The factors are represented in italics such as the obese fraction of the population, management of prediabetes, self-monitoring, etc. Mainly, three interventions were compared and discussed: enhanced clinical management of diabetes, increased management of prediabetes, and reduced obesity prevalence. The simulation results showed that increased clinical management of diabetes was able to lower the prevalence of death, but due to a “backing up” effect, the prevalence of diabetes increased. The intervention focusing on increased management of prediabetes helped to reduce the prevalence of diabetes and death, but could not halt the increases of both events. By contrast, it was found that an intervention that reduces obesity prevalence could reduce the prevalence of both diabetes and death and meanwhile stops the growth in the prevalence of diabetes and death (13).

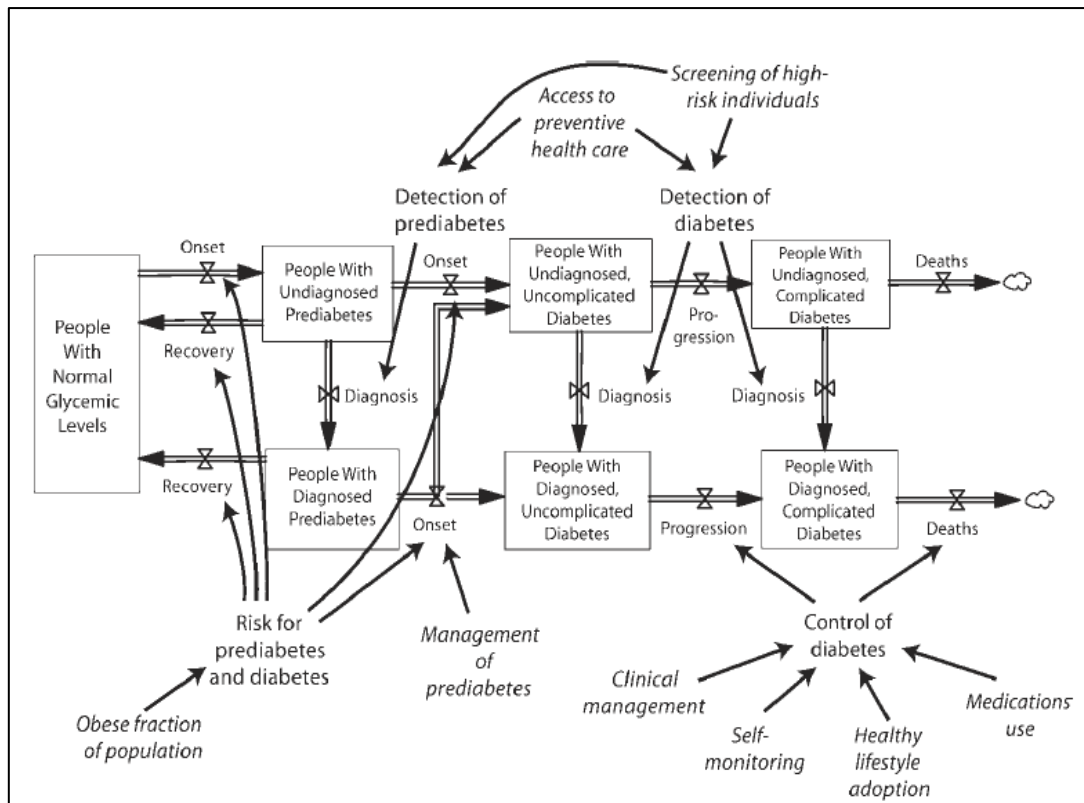


Figure 2.5 Overview of model structure with population stocks and flows and policy related intervention (shown in italics) (13).

SDM also has been used as a tool to evaluate objectives of health care policies (14). Simulation results suggested that a stated objective of 38% reduction in the prevalence of diagnosed diabetes in the United States in *Healthy People 2010* is unlikely to be met, and diabetes prevalence is more likely to continue increasing since there are more new diabetes cases than deaths of diabetes. Moreover, there are conflicts among objectives. For example, the objective to increase detection rate of diabetes and to reduce the diabetic mortality rate are in conflict with the objective to lower the prevalence of diabetes. The first two objectives would actually increase the prevalence of diabetes.

Furthermore, SDM was used to make prediction of the ESRD population in Japan

in 1991 (12). There were two scenarios tested using the validated model: a successful increase or a failure to increase the rate of kidney transplantation. The model's simulation results showed that with no increase of kidney transplantation rate, the number of ESRD patients would be 157,350 while the number would drop to 145,750 if the increase in the rate of kidney transplantation is successful. Moreover, under the second scenario, annual medical expenditure from dialysis of ESRD patients would drop by about 38.6 billion yen.

ABM is useful to help in understanding the behaviors of autonomous agents and the emergent performance of the system in which these individuals circulate (10, 63-66). Hammond (64) discussed using ABM to study the complex dynamics of obesity epidemic and to design policy and intervention. In this study, characteristics of epidemic of obesity were reviewed, such as wide range of scale levels, huge diversity of factors that could affect energy balance, different mechanisms, and interactions and feedbacks among mechanisms. ABM was proposed to be a useful methodology to study the complex dynamics of obesity by addressing attributes of obesity epidemic. Moreover, ABM could also be useful to act as a computational laboratory assisting in understanding “what if” and test different intervention policies (64, 65).

ABM has also been used in the context of infectious disease. Figure 2.6 presents one ABM model for tuberculosis (TB) transmission with smoking as a risk factor (66). Two parts related to one person's status were captured. One is status of TB, such as uninfected state, latent state and active TB state. The other classification captured related to smoking. One SDM model was also built for TB transmission with smoking

being a risk factor. Results from ABM and SDM were compared in this research work, and it was concluded that SDM is a continuous deterministic model and gives only one trajectory of outcome, but stochastic ABM captures stochastic variability in evolution, and – by running a set of Monte Carlo simulations – can be employed to capture the range of outcomes that might be expected. Moreover, it has been discussed that SDM assumes that population is well mixed and is infeasible to incorporate large amounts of heterogeneity, network topologies, and history information, all of which are readily incorporated in ABM models.

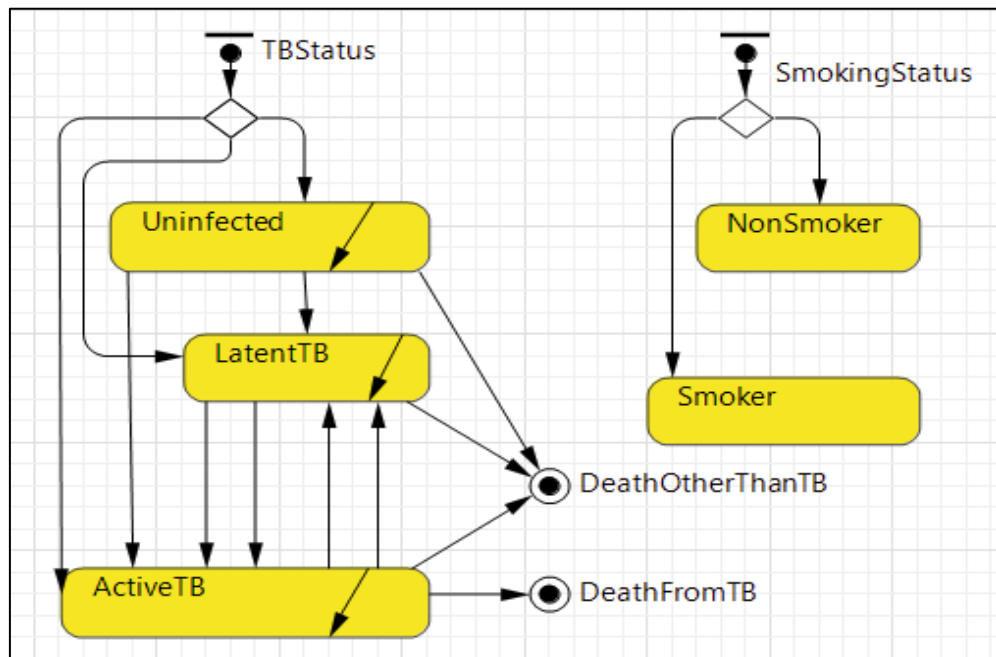


Figure 2.6 Structure of agent-based model of TB diffusion with the effect of smoking (66)

In summary, both SDM and ABM methods have been used in the context of public health to understand complex dynamics of health system and evaluate policy interventions. Both methods will be used in the context of competing risks of ESRD and death without ESRD. The basic components and definition of SDM and ABM

will be discussed in the Methodology sections 5.3 and 5.4 of this thesis. A Comparison of the methods in the context of this work is included in the Discussion section.

### **Chapter 3 Rationale**

ESRD disproportionately affects Aboriginal people in Canada. The reasons for the greater rates of diabetic ESRD among FN with diabetes compared to OSK with diabetes may include an increased onset rate of diabetic glomerulosclerosis that may result in ESRD (7); higher rates of kidney function loss after onset of CKD (7); and/or reduced mortality during pre-ESRD stages of CKD (7). We briefly expand on these hypotheses below.

First, there is an increased prevalence of micro- and macro-albuminuria among FN people when compared to other Canadians. In a study of diabetic FN by Hanley et al. (67), FN people experience higher rates of micro- and macro-albuminuria compared with non-FN people. Increased rates of diabetic glomerulosclerosis may lead to higher rates of ESRD (7).

Second, increased speed of kidney function loss may lead to a higher prevalence of severe CKD among FN than non-FN people, leading to higher rates of ESRD. If kidney function is lost rapidly, the prevalence of pre- and stage 1 CKD as well as later stages of CKD before ESRD should be high (8, 51). Although Gao et al. (8) reported an overall lower prevalence of CKD among Aboriginal than non-Aboriginal people, the prevalence of late-stage CKD among First Nations people was almost twice as high as that among non-First Nations people. This suggests that rapid kidney function loss may lead to increased rates of ESRD.

Third, FN people may have enhanced survival before reaching the stage of ESRD,



resulting in a higher rate of ESRD development. Death without ESRD is a competing risk event for ESRD cases and is common among people who have CKD (68). One study in Saskatchewan by Dyck et al. (7) found that diabetic FN with early stage CKD ( $\text{eGFR} \geq 30 \text{ ml/min}$ ) experienced a older age- and sex-adjusted risk of death compared with diabetic OSK. However, mortality risks for people at the pre-ESRD stage of CKD, with and without adjusting for age and sex, were lower among FN than OSK. This is likely to reflect the fact that FN people typically acquire diabetes at a significantly younger age than do non-FN people.

This study will be the first, to our knowledge, that examines disparities in the risks of ESRD and mortality and estimates the simultaneous risks of these two events between FN and OSK using competing risks survival analysis in the population of Saskatchewan. Simulation modeling is useful for evaluating the accuracy of the hazard rate estimations. A better understanding of these disparities will help to target research, intervention and disease management to decrease the ethnic disparities, such as applying the knowledge to more effective diabetes and CKD prevention and CKD management initiatives.

## **Chapter 4 Study objectives**

1. To increase our understanding of why there are disparities in ESRD rates between FN and OSK with diabetes.
  - a. To determine whether there are significant disparities in the risk of ESRD development between FN and OSK.
  - b. To determine whether there are significant disparities in the risk of mortality without ESRD between FN and OSK.
2. To build Saskatchewan diabetic ESRD dynamic models.
  - a. To estimate the hazard rates for ESRD and mortality without ESRD and to build a dynamic model using hazard rate estimations.
  - b. To validate models by comparing model-predicted data with historical data.

## **Chapter 5 Methodology**

### **5.1 Study design and participants**

Administrative databases were used to collect cohorts of First Nations (FN) people and other Saskatchewan (OSK) people with diabetes and ESRD between 1980 and 2005. This study received approval from the University of Saskatchewan Ethics Review Board. The detailed description of the study populations can also be found in two publications by Dyck et al. (6, 23).

The total population was then divided into FN and OSK as follows. FN people in this study are people who are indigenous to Canada. FN people are registered under Section 6 of the Indian Act of Canada and are beneficiaries of universal health care. OSK in this study population are a group of people who have predominantly European origin but also include non-registered FN (<0.5%) and Métis (approximately 5%) (6, 23).

The diabetes case definition required one hospital discharge for diabetes (Hospital Service database), two physician service claims from the Physician's Service database, or a physician service claim and then a hospital discharge for diabetes within 730 days. We excluded gestational diabetes cases from the study population and only included type 1 and type 2 diabetes. The ESRD case definition was based on physician service codes for chronic dialysis and also included renal transplantation. In addition, chronic dialysis was defined as 90 days' dialysis with no break for more than 21 days. In the

study population of diabetics, the causes of ESRD were not distinguished by etiology (6, 23).

After defining people who had diabetes and ESRD, we then obtained supplementary information, such as sex, ethnicity, birth year, year at diagnosis of diabetes, and year at ESRD diagnosis, as well as year of death and loss of health care coverage if those occurred.

In our study, people who developed ESRD before diabetes were excluded from the study population, but they remained if they were diagnosed with both diseases in the same year. Death without ESRD was treated as a competing risk event with ESRD, but death after the development of ESRD or in the same year as ESRD development was not considered a competing risk event. Moreover, we excluded people who were younger than 20 years old at the time of their diabetes diagnosis so that we could exclude most type 1 diabetes cases.

## **5.2 Statistical analysis**

### **5.2.1 Introduction**

Descriptive statistics were used to gain information about distributions of variables, such as diabetes diagnosis age and ethnicity. Comparisons between groups were performed using t-tests and  $\chi^2$  tests, The significance level was set to 0.05. Competing risks survival models were used to model the relationship between predictors and outcomes. Cox cause-specific model, Weibull proportional hazards model and piece-wise exponential proportional hazards model were used. Predictors in our study

included diabetes diagnosis age, sex and ethnicity, and there were two outcomes: ESRD and death without ESRD.

Survival analysis is “an application of a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs” (69). One straightforward method for survival analysis is a non-parametric method, such as the life table method introduced by Berkson and Gage (70). Another non-parametric method was proposed by Kaplan and Meier (71). Non-parametric methods can be used to analyze one sample of survival data or to compare two or more groups of survival data. Moreover, supplementary information can be recorded for individuals, such as age, sex and ethnicity. These variables can be referred to as explanatory variables. Non-parametric methods cannot be used to determine the effects of explanatory variables on survival times. The Cox proportional hazards (PH) model can be used to analyze survival data with explanatory variables (72). Sometimes, if the data follow a particular probability distribution – such as the exponential distribution, Weibull distribution or Gompertz distribution – we can use parametric survival analysis to test hypotheses and also to estimate hazard rates at the individual level, given certain characteristics of individuals.

We first give the following definitions of an event, survival time and censoring (69). An “event” denotes a death or disease incidence that may happen to an individual in a certain study. The “survival time” means the number of time units (e.g., years, months, weeks, or days) from the beginning of the follow-up period until an event occurs. In our study, time means years since diabetes diagnosis. Censoring happens when part of an

individual's survival time is partially known, but we do not know the whole survival time exactly. Here, we will discuss the right censoring that happened in our study (69). In each of these situations, a patient who entered a study at time  $t_0$  experienced an event of interest at time  $t_0+t$ . However,  $t$  was unknown, either because the individual was still alive or because he or she had been lost to follow-up. If the individual was last known to be alive at time  $t_0+t_c$ , the time  $t_c$  is called a censored survival time. Right censoring occurs when censored survival time is less than the actual but unknown survival time (69).

### **Survivor Function**

Two functions can be used to describe survival data: the survivor function and the hazard function (69, 72). Variable  $T$  is a random variable denoting the actual survival time  $t$ , ( $T>0$ ). The probability density function for  $T$  is  $f(t)$ . The cumulative distribution of  $T$  is given by

$$F(t) = P(T < t) = \int_0^t f(u)du \quad (5.1)$$

and represents the probability that the survival time is less than time  $t$ . The survivor function  $S(t)$  is defined as the probability that the survival time is greater than or equal to  $t$ , given by

$$S(t) = P(T \geq t) = 1 - F(t) \quad (5.2)$$

and therefore is used to represent the probability that an individual survives from the start time until a time equal to or beyond  $t$ .

A hazard function expresses the hazard risk of an event at some time  $t$ , conditionally that person has survived until time  $t$ . The formula for  $h(t)$  is shown below:

$$h(t) = \lim_{\delta t \rightarrow 0} \left\{ \frac{P(t \leq T < t + \delta t | T \geq t)}{\delta t} \right\} \quad (5.3)$$

$H(t) = \int_0^t h(u) du$  is called the cumulative hazard function.

### 5.2.2 Non-parametric methods

An initial step in the analysis of survival data is to present numerical or graphical summaries of the survival time. Non-parametric methods can be used to estimate survival data through estimations of the survivor function and hazard function. This method does not require any specific assumptions regarding the distribution of the survival time (72).

#### The Kaplan-Meier estimate of the survival function

To obtain a Kaplan-Meier (K-M) estimate, we first create a series of time intervals based on the time of events (72). In each interval, there is only one event that happens at the beginning of the interval. Suppose that there are  $n$  individuals with observed survival times  $t_1, t_2, \dots, t_n$  during this period of time, with some of these observations being right censored. There are  $r$  event times, and  $r$  is smaller than or equal to  $n$ , and after arranging time in ascending order, the  $j$ th time interval is denoted  $t_j$ , for  $j=1, 2, \dots, r$ , and so the  $r$  ordered times are  $t_1 < t_2 < \dots < t_r$ . The number of people who are alive just before time  $t_j$  is  $n_j$ , for  $j=1, 2, \dots, r$ , and  $d_j$  will denote the number of individuals who

experience the event at time  $t_j$ . The time interval from  $t_j - \delta$  to  $t_j$ , where  $\delta$  is an infinitesimal time, includes only one event. The probability for an individual to experience the event during this interval is estimated to be  $d_j/n_j$ . Therefore, the probability for an individual to survive through that interval will be  $(n_j-d_j)/n_j$ . Therefore, the K-M estimator of  $S(t)$  for  $t_k \leq t < t_{k+1}$ ,  $k=1,2,\dots,r$  is defined as:

$$S(t) = \prod_{j=1}^k \left( \frac{n_j - d_j}{n_j} \right) \quad (5.4)$$

representing the probability of surviving through interval  $t_k$  to  $t_{k+1}$  and all preceding intervals (72).

### Hazard Function

Smoothed plots of hazard rates are used often in practice, and they can give us an understanding of the shape of hazard curves, thereby allowing us to assume the distribution of survival time. The Epanechnikov kernel smoothed estimate of hazard function is used (72, 73), and it is based on  $r$  ordered death times,  $t_{(1)}, t_{(2)}, \dots, t_{(r)}$ ,

$$h^*(t) = b^{-1} \sum_{j=1}^r 0.75 \left\{ 1 - \left( \frac{t - t_{(j)}}{b} \right)^2 \right\} \frac{d_j}{n_j} \quad (5.5)$$

with  $d_j$  deaths and  $n_j$  at risk at time  $t_{(j)}$ , and  $t$  is defined on the time interval  $(b, t(r)-b)$ . The parameter  $b$  is named bandwidth, and we can obtain an optimal estimation for it. For any time point  $t$ , death time in interval  $(t-b, t+b)$  is used to estimate hazard rate at time  $t$ .



### 5.2.3 Cox proportional hazards model

The non-parametric methods described above can be used to describe one group of survival data or compare survival time between two groups – for example, a comparison of survival time between FN and OSK following diabetes diagnosis. However, there are situations in which more information about patients – referred to as explanatory variables in the dataset, such as age, sex, ethnicity, and blood pressure – can be obtained. Researchers often wish to determine whether there are relationships between these explanatory variables and survival time. To accomplish these aims, statistical modeling procedures should be used. The Cox proportional hazards (PH) model (72), which is commonly used to explore how explanatory variables impact hazard ratios. Suppose that the event hazard at a specific time depends on the values  $x_1, x_2, \dots, x_p$  of  $p$  explanatory variables  $X_1, X_2, \dots, X_p$ . Let  $h_i(t)$  be the hazard function for the  $i$ th individual at time  $t$ . We call the hazard function for a person for whom the values of all the explanatory variables are zero the baseline hazard function, denoted as  $h_0(t)$ .

The Cox PH model can be expressed for the  $i$ th person at time  $t$  as follows:

$$h_i(t) = h_0(t) \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) \quad (5.6)$$

where  $x_{1i}, x_{2i}, \dots, x_{pi}$  are the values of  $p$  explanatory variables for the  $i$ th person, and  $\beta_1, \beta_2, \dots, \beta_p$  are the coefficients of the corresponding explanatory variables.

## 5.2.4 Parametric proportional hazards model

In the Cox cause-specific model, there is no assumption required regarding the distribution of survival time. If a certain distribution for survival time is valid, we may use the parametric model to test the effect of explanatory variables and estimate hazard rates with explanatory variables information included in the estimation of hazard rates. The Weibull PH model and exponential PH model will be discussed below.

### 5.2.4.1 Weibull proportional hazards model

Under the Weibull PH model, the hazard of an event can be expressed with scale parameter  $\lambda$  and shape parameter  $\gamma$ , ( $\lambda > 0$ ,  $\gamma > 0$ ), and the hazard function of one  $W(\lambda, \gamma)$  distribution is given by  $h(t) = \lambda \gamma t^{\gamma-1}$ . Under the Weibull PH model, the hazard function of the  $i$ th person at time  $t$  with covariates  $(x_{1i}, x_{2i}, \dots, x_{pi})$  can be shown as:

$$h_i(t) = \lambda \gamma (t)^{\gamma-1} \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) = \lambda \gamma (t)^{\gamma-1} \exp(\beta' \overrightarrow{X_i}) \quad (5.7)$$

$\beta' \overrightarrow{X_i}$  indicates  $\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}$ . Therefore, we can say that the survival time of the  $i$ th person at time  $t$  follows Weibull distribution with scale  $\lambda \exp(\beta' \overrightarrow{X_i})$  and shape parameter  $\gamma$ .

$$\lambda = \exp(-\mu / \sigma) \quad (5.8)$$

$$\gamma = 1 / \sigma \quad (5.9)$$

$$\beta_m = -\alpha_m \quad (5.10)$$

$m=1, 2, \dots, p$  and  $\mu, \gamma, \alpha_m$  can be estimated from statistical analysis procedures.

#### 5.2.4.2 Exponential PH model

The exponential PH model (72) assumes that the hazard function is constant over time (72). The hazard function under this model can be written as  $h(t)=\lambda$ , for  $0 \leq t < \infty$ .  $\lambda$  can be estimated by fitting the model to the observed data, and it is a positive constant.

Under the exponential proportional hazards model, the hazard function of a particular  $i$ th patient at time  $t$  may be given as:

$$h_i(t) = \lambda \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) \quad (5.11)$$

With  $\beta_1, \beta_2, \dots, \beta_p$  being coefficients of explanatory variables, with

$$\lambda = \exp(-\mu) \quad (5.12)$$

$$\beta_m = -\alpha_m \quad (5.13)$$

and with  $m=1, 2, \dots, p$ , similar to the Weibull PH model, we can obtain the estimates of  $\mu, \alpha_m$  from statistical analysis.

#### 5.2.4.3 Piece-wise exponential proportional hazards (PH) model

The piece-wise exponential PH model (74) is an extension of the exponential proportional hazards model. For constructing a piece-wise exponential model, we first need to specify a time grid  $\tau = \{s_0, s_1, s_2, \dots, s_k\}$  and  $0=s_0 < s_1 < s_2 < \dots < s_k < \infty$ . Then, the time axis is divided into  $p$  intervals  $I_j = (s_{j-1}, s_j]$ , for  $j=1, \dots, k$ . Then, we assume that the hazard rate during each interval of the grid  $\tau$  is constant, which means that for  $t \in I_j$ ,  $j=1, \dots, k$ :

$$h(t) = \lambda_j \quad (5.14)$$

Under the piece-wise exponential proportional hazards model, the hazard function of a particular  $i$ th patient at time  $t$  for time interval  $I_j$ , where  $t \in I_j$ , and  $j=1, \dots, k$  may be given:

$$h_i(t) = \lambda_j \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}) \quad (5.15)$$

where  $\beta_1, \beta_2, \dots, \beta_p$  are coefficients of exploratory variables. We can determine  $\mu, \alpha_m$  from statistical analysis, with  $m=1, 2, \dots, p$

$$\beta_m = -\alpha_m \quad (5.16)$$

To compare parametric models, we can use statistical tests such as the likelihood ratio (72) test or Akaike's information criterion (AIC). The likelihood ratio test can be used to compare nested models such as the exponential distribution, Weibull distribution and gamma distribution, while AIC can be used to compare models that are not nested. AIC is defined as

$$AIC = -2 \log L + \alpha q \quad (5.17)$$

where  $L$  is the estimated maximized likelihood for one model,  $q$  is the number of explanatory variables, and  $\alpha$  is a predetermined constant value. After fitting one model to survival data, we can obtain an AIC value for the fitted model. The smaller AIC is, the better the model was fitted.

### 5.2.5 Competing risks survival analysis

Competing risks occur when a person experiences a chance of two or more events (69). For example, Figure 5.1 presents one situation of competing risks. One diabetic patient may die before they develop ESRD; therefore, death without ESRD and ESRD

are competing risk events for people with diabetes. One person with diabetes can only experience either of these possible events at one time (75).

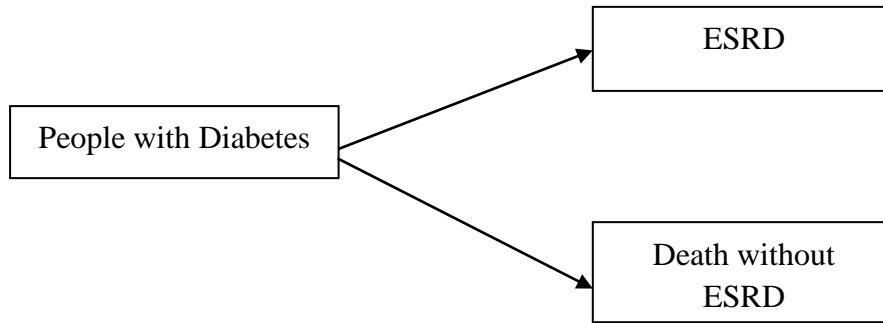


Figure 5.1 Competing risk events.

#### 5.2.5.1 Cause-specific hazard function

First, we give the definition of a cause-specific hazard function as:

$$h_c(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T_c < t + \Delta t, C = c | T_c \geq t)}{\Delta t} \quad (5.18)$$

$T_c$  is time to failure from event  $c$ ,  $c=1, 2, \dots, C$ , where  $C$  is the number of event types (69).

#### 5.2.5.2 Cumulative incidence function

Another nonparametric method is to use a cumulative incidence function (CIF), which indicates the probability of certain event (e.g., ESRD or death without ESRD) before some point in time. To calculate CIF, we need several estimations from the data, including the Kaplan-Meier estimator and cause-specific hazard. First we order the event

time  $t_1 < t_2 < \dots < t_r$  and make  $d_{cj}$  the number of events of type  $c$  at time  $t_j$ . At time  $t_j$ , the number of people at risk is  $n_j$ .  $S(t)$  is the Kaplan-Meier estimator of the probability of being event free by time  $t$ , and  $h_{cj}$  is the cause-specific hazard for event  $c$  at time  $j$ . CIF for event type  $c$  can be given by (69):

$$F_c(t) = \sum_{j, t_j \leq t} h_{cj} S(t_{j-1}) \quad (5.19)$$

The cause-specific hazard at time  $t_j$  can be expressed as  $d_{cj}/n_j$ . Therefore,

$$F_c(t) = \sum_{j, t_j \leq t} \frac{d_{cj}}{n_j} S(t_{j-1}) \quad (5.20)$$

### 5.2.5.3 Cox cause-specific model

The Cox cause-specific hazard model is used to analyze competing risk survival with a Cox proportional hazards model to estimate hazard ratios between groups for each type of failure, treating the other failures as censoring together with withdrawal from the study or loss to follow-up (69).

The Cox cause-specific model for event type  $c$  with covariates for the  $i$ th patient at time  $t$  with  $p$  covariates  $\mathbf{X}=(x_1, x_2, \dots, x_p)$  is defined to be:

$$h_c(t, \mathbf{X}) = h_{0c}(t) \exp \left[ \sum_{i=1}^p \beta_{ic} x_i \right] \quad (5.21),$$

$c=1, 2, \dots, C$ , where  $C$  is the number of event types.

The Cox cause-specific hazard model is used to compare the hazard ratios of developing ESRD or mortality between groups such as FN and OSK.

### 5.3 System Dynamics modeling

After fitting the statistical model to the data set and getting estimates of hazard rates, we want to build up a SDM using the subscribing approach to represent the Saskatchewan Diabetes and ESRD system. There are several advantages of SDM, such as helping to capture the complexity of health system by including many aspects of health problems, using prospective simulations to anticipate possible future trends, and helping policy makers to evaluate policies by performing what-if experiments within the model (76). SDM is most commonly conducted at the aggregate level of modeling, where it depicts a series of cross-sectional views of population evolution.

#### 5.3.1 System Dynamics modeling, first order delay and competing risks

Stocks and flows are central concepts in System Dynamics. This section first introduces these two terms with a diagram and then discusses mathematical representation of stocks and flows (76). Figure 5.2 depicts an example stock and flow diagram.

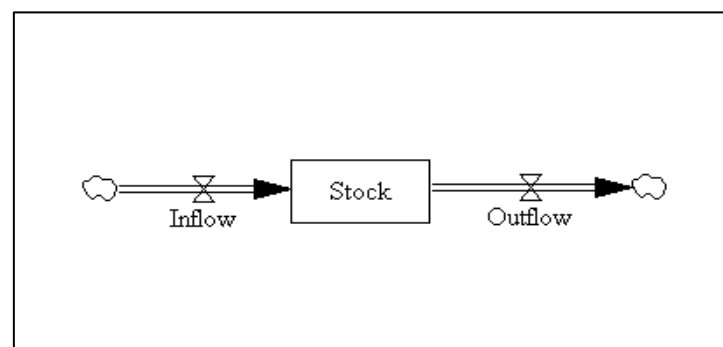


Figure 5.2 Stock and flow

Stocks are indicated by rectangles and flows are indicated by arrows. Inflows for a

given stock are shown as going into that stock and outflows are shown as going out of that stock. Valves on flows control the flows. Clouds can be used to indicate the source for an inflow or an outflow's destination. One can regard the stock as a bathtub with the water level representing the value of the stock, such as prevalence of diabetes. The rate of inflow can be analogous to the rate of water flowing into the bathtub, while the rate of outflow is analogous to the rate of water flow out of the bathtub (76). Stocks and flows have mathematical meaning. Stocks integrate the net flow. The structure represented in Figure 5.2 can be understood in a mathematical context with the following integral equation (76):

$$Stock(t) = \int_{t_0}^t [Inflow(s) - Outflow(s)]ds + Stock(t_0) \quad (5.22)$$

where  $Stock(t)$  represents stock value at time  $t$ , and  $Inflows(s)$  and  $Outflows(s)$  indicate the values of inflows and outflows at time  $s$  between initial time  $t_0$  and time  $t$ .

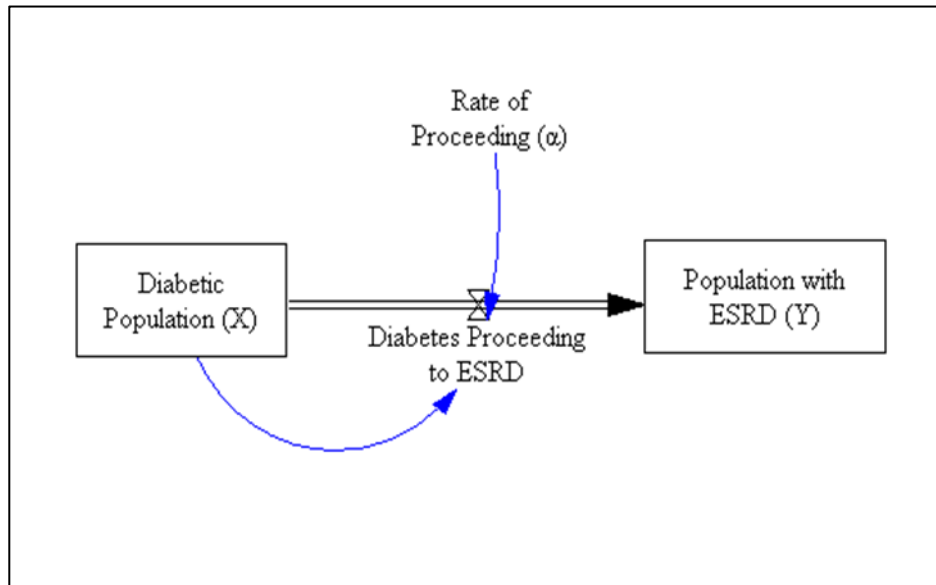


Figure 5.3 First order delay for progression from diabetes to ESRD



In SDM, first order delay is a key characteristics and it could be treated as a representation of deterministic approximation to a memory-less stochastic process (15). The system is called “memory-less” because the chance of a person leaving in the next unit of time is independent of how long they have been in the state. A first order delay for diabetic population progressing to ESRD or death without ESRD is shown with stock and flow charts in Figure 5.3. The following derivation used Dr Osgood’s class slide as reference (15). Alpha ( $\alpha$ ) is the hazard of progressing from “Diabetic Population ( $X$ )” to “Population with ESRD ( $Y$ )”, indicating the probability of progressing from diabetes to ESRD per unit time. While  $\alpha$  commonly changes in System Dynamics models (e.g. for the force of infection associated with epidemic spread), for the discussion below, we assume that alpha is a constant. The total change in the value of the stock over an infinitesimally small time duration  $dt$  is  $dx$ . Thus,

$$dx = -x\alpha dt \quad (5.23)$$

$$\frac{dx}{x} = -\alpha dt \quad (5.24)$$

The value of stock  $x$  at time 0 is the initial value indicated as  $x(0)$ , and the value of  $x$  at time  $T$  is  $x(t)$ .

$$\int_{t=0}^{t=T} \frac{dx}{x} = \int_{t=0}^{t=T} -\alpha dt \quad (5.25)$$

$$\frac{x(T)}{x(0)} = e^{-\alpha T} \quad (5.26)$$

indicating the fraction of the initial population remaining in the stock at time  $T$ . The probability of a given person leaving in a time  $dt$  conditional on that person having remained in the stock up to that time  $t$  is always equal to  $\alpha dt$ . Hazard rate estimations using the piece-wise exponential PH model in Section 5.2.4.3 could be used as parameters for rates of proceedings.

The likelihood that someone leaves between time  $t$  and  $t+dt$  is  $P(\text{leaving exact between time } t \text{ and } t+dt) = P(\text{person remains in the stock at time } t) * P(\text{leaving exactly between time } t \text{ and } t+dt | \text{ person remains in the stock at time } t) = \alpha e^{-\alpha t} dt$ .

The mean time for people to stay in the stock can be estimated by:

$$\int_{-\infty}^{\infty} tp(t)dt = \int_0^{\infty} tp(t)dt = \int_0^{\infty} t\alpha e^{-\alpha t} dt = \alpha \int_0^{\infty} te^{-\alpha t} dt = 1 / \alpha$$

Competing risks of ESRD and mortality are shown with stocks and flows (15) in Figure 5.4. There are two outflows from the “Diabetic Population” stock. One flow named “Diabetes progressing to ESRD” was directed to the stock of “Population with ESRD” and the other, named “Deaths of Diabetic Population”, was also flowing out of the stock. Suppose a diabetic population’s initial value is  $x(0)$  at time zero, and its value at time  $T$  is  $x(T)$ . In the model, a likelihood density named “Rate of progressing from Diabetes to ESRD ( $\alpha$ )” was used to determine the value of the flow “Diabetes progressing to ESRD” and the other likelihood density (named “Annual Risk of Diabetic Mortality ( $\beta$ )”) was used to determine the flow “Death of Diabetic Population”.

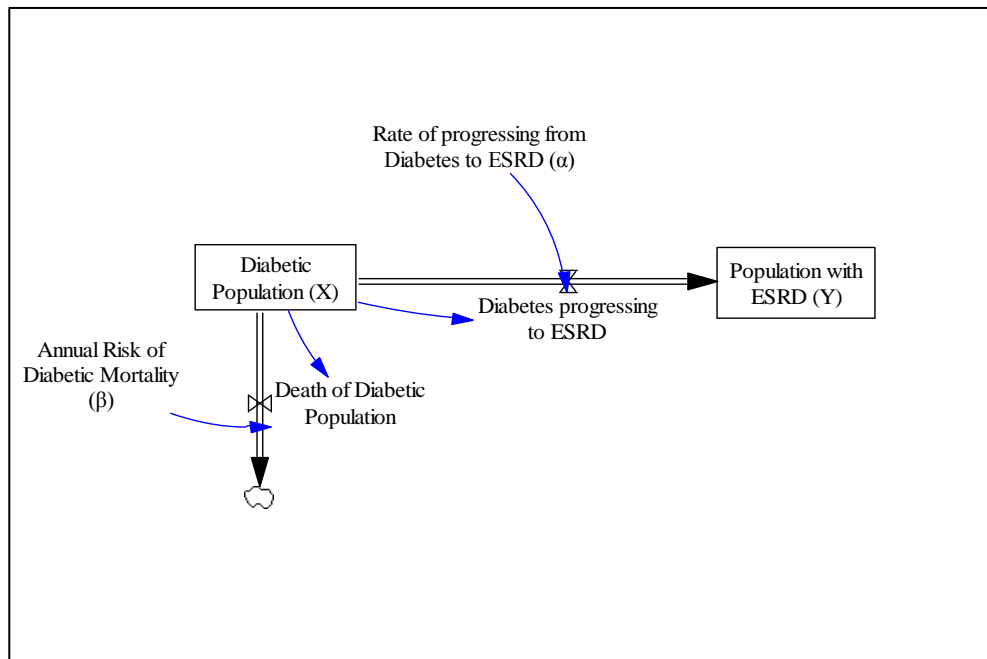


Figure 5.4 Competing risks between ESRD and death without ESRD.

The total change in the value of the stock  $x$  over an infinitesimally small time duration  $dt$  is  $dx$ . The following derivation used Dr. Osgood's class slides as reference (15).

$$\frac{dx}{dt} = -\alpha x - \beta x = -(\alpha + \beta)x \quad (5.27)$$

Suppose that at time zero, the initial value of the Diabetic Population is  $x(0)$ , and we want to find the value of  $x$  at time  $t$ .

$$x(t) = x(0)e^{-(\alpha+\beta)t} \quad (5.28)$$

For  $t > 0$ ,  $P(\text{Person remains in the stock at time } t) = e^{-(\alpha+\beta)t}$

$P(\text{leaving exactly from diabetes stock to ESRD stock between time } t \text{ and } t+dt|$

$\text{Person remains in the stock at time } t) = \alpha * dt$

The likelihood  $A(t)$  that someone leaves from the diabetes stock to the ESRD stock between time  $t$  and  $t+dt$  is  $P(\text{leaving exact between time } t \text{ and } t+dt) = P(\text{Person remains in the stock at time } t) * P(\text{leaving exactly between time } t \text{ and } t+dt | \text{ Person remains in the stock at time } t) = \alpha e^{-(\alpha+\beta)t} dt$

Here  $\alpha$  and  $\beta$  are constant. Values for hazard rates  $\alpha$  and  $\beta$  can be estimated from statistical procedures modeling the likelihood that diabetic people will develop ESRD and die, respectively (15).

### 5.3.2 Subscripting Dynamics Modeling

Sometimes during modeling, one piece of a model structure can be repeated several times (77). For example, if we divide age at diabetes diagnosis into several age categories, the process of developing ESRD will be replicated for different diagnosis age categories. One method for summarizing this repeating structure is to create one structure and then to replicate that structure several times as needed both visually and functionally. Visually, this will lead to complex graphs of stocks and flows, which can be confusing to interpret. Even more importantly, changes in that process that apply across all of the different instances of that structure will typically need to be made manually to each section in turn, in a manner that is tedious and error prone.

Another way to construct a repeated structure is to use subscripts. We can create one original structure and add other structures through the addition of subscript elements. Values for each subscript can be changed, and the overall diagrams will look clearer. Because of the centralized description of the system across all subscripts,

changes can be localized to one point. For example, in the model examined here, we need to divide ethnicity into FN and OSK, and we can therefore use subscripts to organize diagrams as shown in Figure 5.5. Each subscripted structure can be described and given distinct values where required.

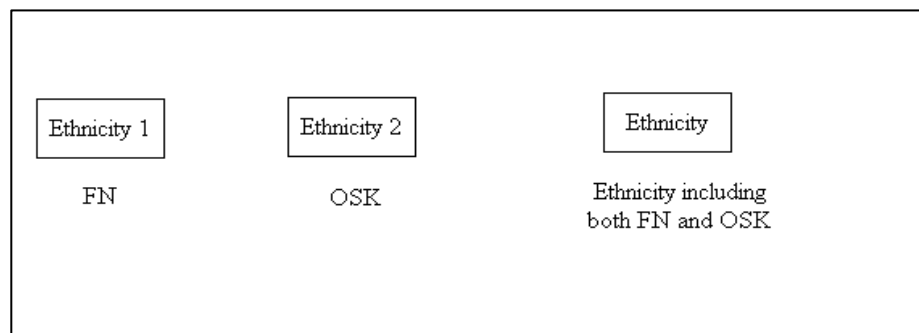


Figure 5.5 Subgroup in Vensim

#### 5.4 Agent-based modeling

“In agent-based modeling (ABM), a system is modeled as a collection of autonomous decision-making entities called agents” (63). Agents are able to adapt and modify behaviors based on certain rules. An agent-based model is made of a system of agents and interactions between them and with an environment. ABM is used to describe and simulate a system, and we can first describe the assumptions of behaviors of agents and the mechanisms of interaction among them. In this thesis, we used Anylogic software as a platform to build and simulate agent-based models. In the following subsection, certain typical terminology and representations in Anylogic will be discussed, such as state, transition, and branch.

### 5.4.1 Statecharts

One statechart is shown in Figure 5.6. It is made of an entry point, branch, state and transition (10, 78):

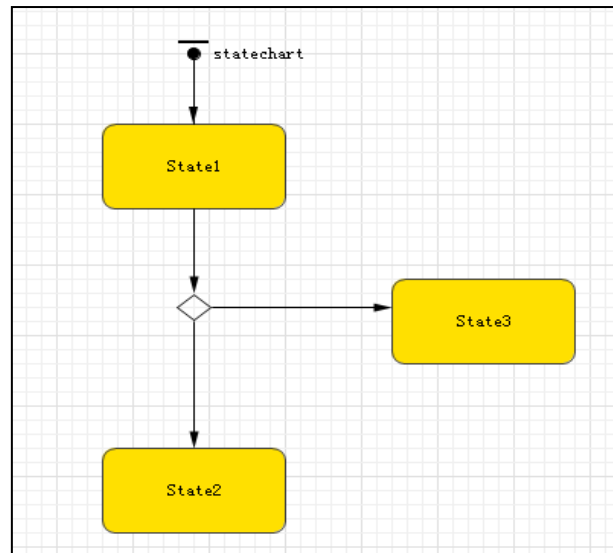


Figure 5.6 Statechart.

An event is used to schedule some action in the model and to model delays and timeouts. A state chart is used in ABM for describing events and behavior during a period of time. States and transitions are two major parts of state charts. Presence in states is controlled by reactions to events. A transition denotes a switch from one state to another. A statechart entry point is used to indicate the initial state of the statechart. There should be exactly one statechart entry point defined for each statechart. “A branch represents a transition branching and/or connection point”(78). Once branches are executed, they create conditional transitions to the next destination states.

ABM is typically stochastic in occurrence of certain events, transitions between states; therefore, different simulations may show different results for the same model

with the same parameters. To derive conclusions that are robust in spite of such variation, the Monte Carlo method was used. After performing Monte Carlo analysis by running one simulation 100 times, we obtain an array of outputs and show them in a 2D histogram. In 2D histogram, we could use envelopes to show percentage of histogram. For example, in specifying envelopes as 0.025, 0.05, 0.25, 0.75, 0.95, 0.975, we are showing areas containing a given 2.5%, 5%, 25%, 75%, 95% and 97.5% of realization results (78).

## **5.5 Software**

Statistical analysis for this thesis was performed using SAS 9.2 (SAS Institute, Cary, NC, USA). Dynamic Models were constructed by using Vensim DSS 5.5c (Ventana Systems, Inc., Harvard, MA, USA) and Anylogic 6.2 Advanced.

## Chapter 6 Results

### 6.1 Description of study population and descriptive analysis

In total, 108,037 people with diabetes were recorded in the health care system from 1980 to 2005. After excluding people who developed diabetes before 1980, people whose diabetes diagnosis age was less than 20 and those that developed ESRD before diabetes, there were a total of 90,429 diabetic people in the study. Figure 6.1 is the data extraction flowchart of the study population. Figure 6.2 and Figure 6.3 present data extraction flowcharts for both ESRD and death without ESRD among FN and OSK, respectively.

We performed descriptive analysis to characterize the variables and their distributions in the data; these results are presented in Table 6.1. We also divided age at diabetes diagnosis into three categories: less than 40, between 40 and 60, and older than 60 years. ESRD age and death age indicate age at ESRD diagnosis and death, respectively. Among the 90,429 diabetic subjects in the study, 8,254 (9%) of them were First Nations people. Among FN, 3,718 (45%) were male. Their mean age at diabetes diagnosis for FN was 47.2 (SD= $\pm 14$ ) years. Among OSK, 44,820 (55%) were male. Their mean age at diabetes diagnosis was 61.6 (SD= $\pm 15.3$ ) years. Of the total FN study population, 200 (2.4%) people developed ESRD, and 1,482 (18%) people died without the development of ESRD. Of the total OSK study population, 600 (0.7%) people developed ESRD, and 28,450 (34.6%) people died without developing ESRD. Mean ages were compared between FN and OSK using the t-test; the results show that the



mean ages at diabetes diagnosis ( $P < 0.0001$ ), at ESRD development ( $P < 0.0001$ ) and at death ( $P < 0.0001$ ) were significantly different between FN and OSK. There was a greater fraction of males among OSK compared with FN ( $P < 0.0001$ ). Compared with OSK, FN people experienced more cases of ESRD and fewer deaths ( $P < 0.0001$ ). Moreover, in the entire dataset, 48 patients were diagnosed with ESRD and diabetes in the same year. For such cases, we recoded the observation time from diabetes diagnosis to ESRD diagnosis as ranging from 0 to 0.5 year. There were 53,627 (59.3%) censoring events (loss to follow-up or end of study) in the study population during the whole study period.

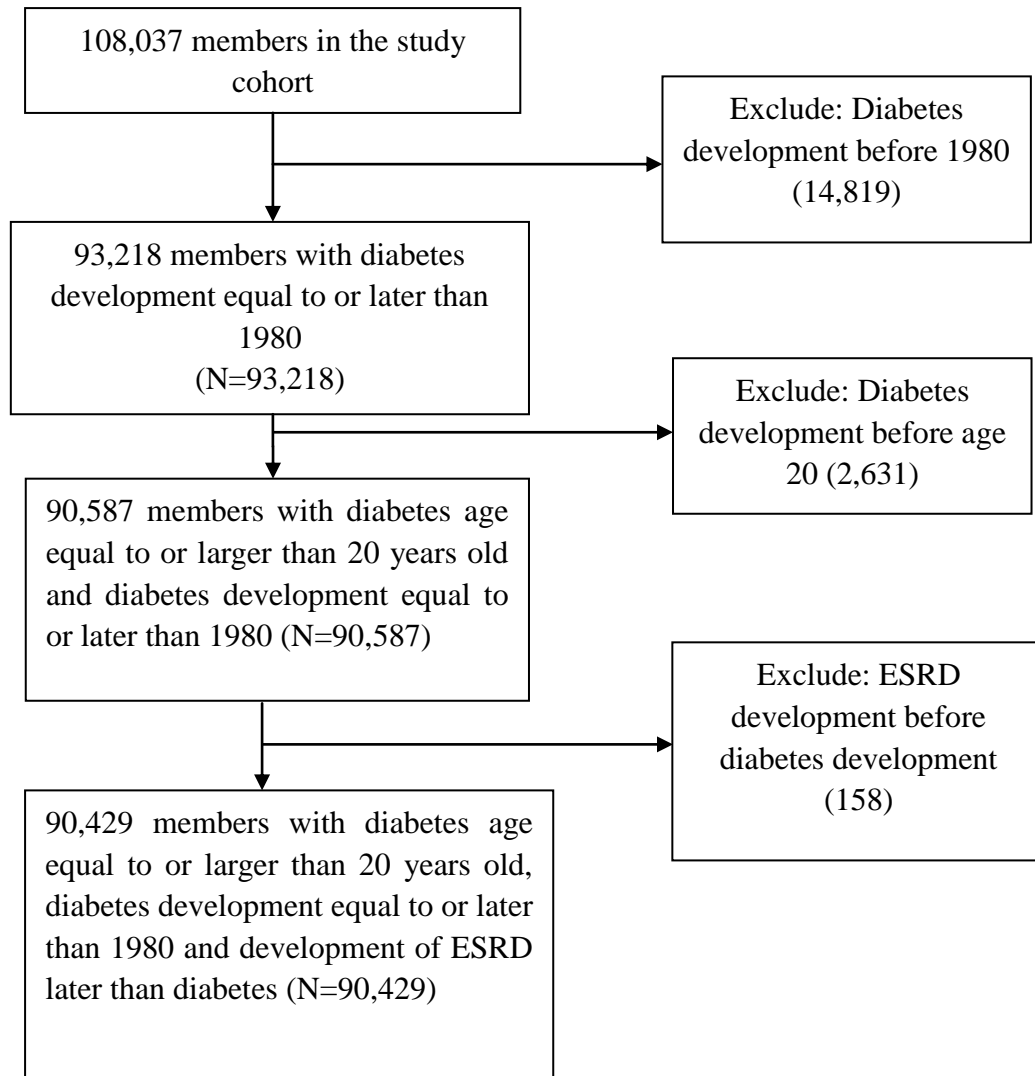


Figure 6.1 Subjects in study.

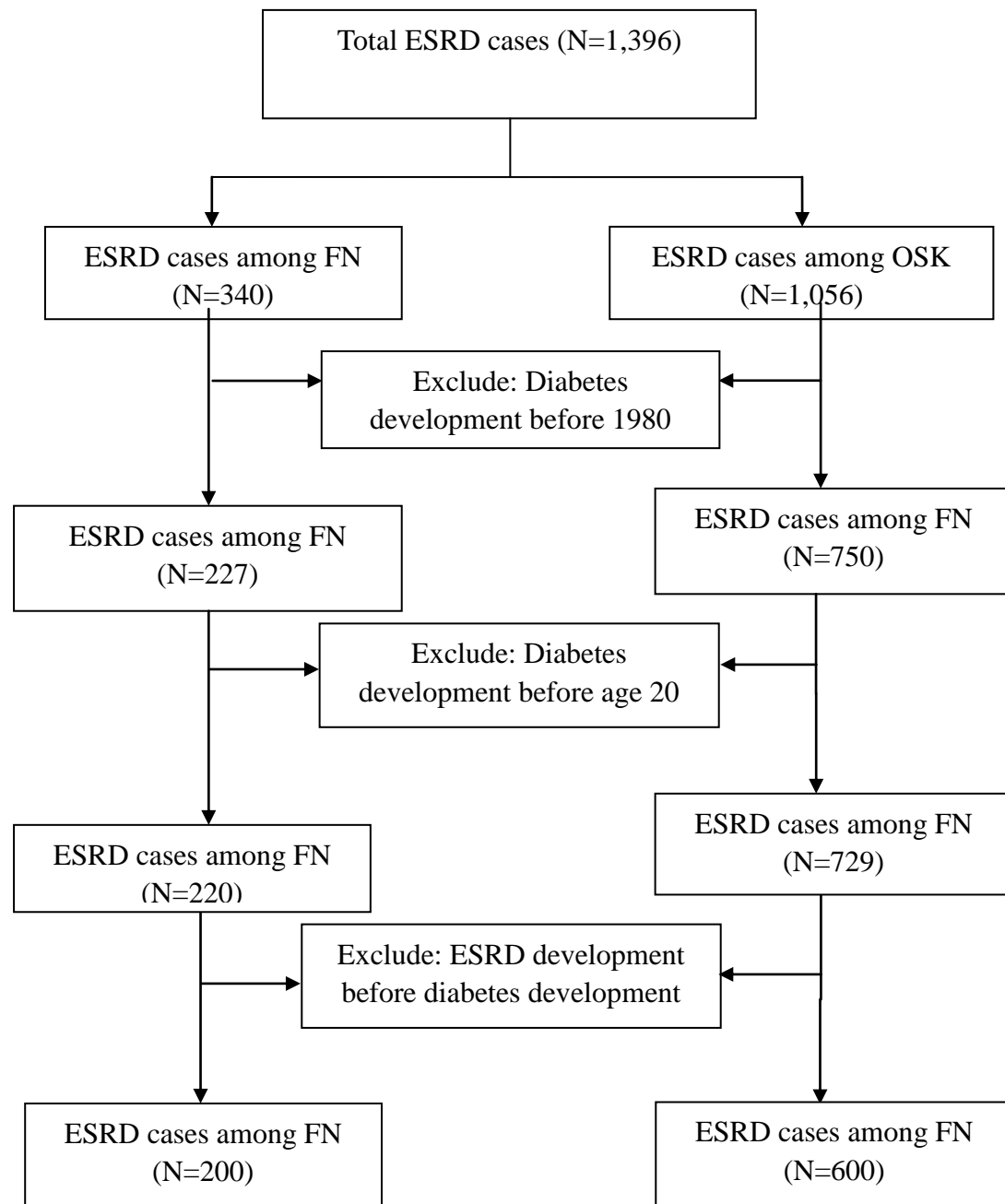


Figure 6.2 ESRD cases by ethnicity.

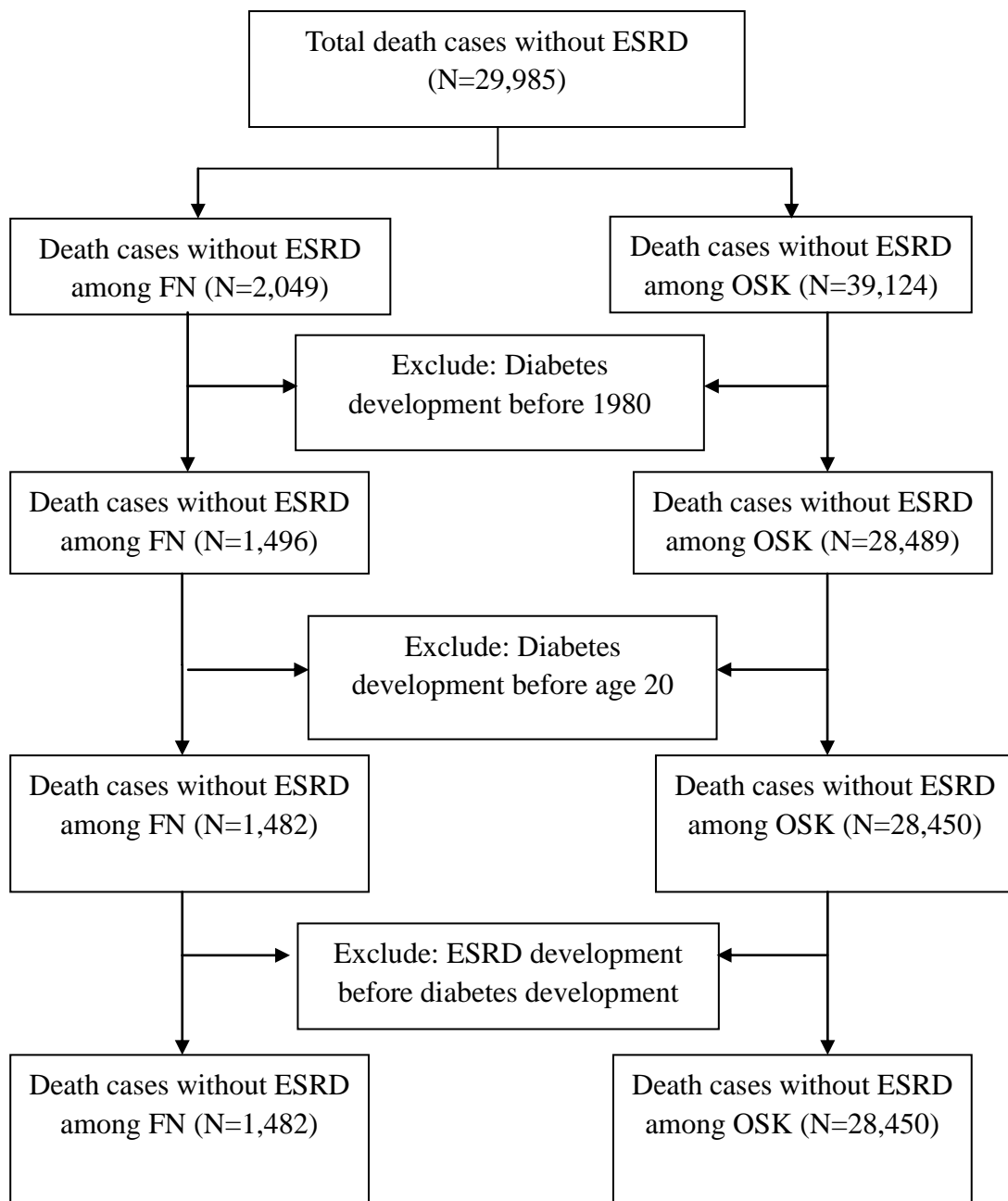


Figure 6.3 Death cases by ethnicity.

Table 6.1. Baseline characteristics by ethnicity.

Variable	FN (N=8,254) Mean ( $\pm$ SD); N (%)	OSK (N=82,175) Mean ( $\pm$ SD); N (%)	P value
Male	3,718 (45.04%)	44,820 (54.5%)	P<0.0001
Diabetes age	47.2 ( $\pm$ 14)	61.6 ( $\pm$ 15.3)	P<0.0001
Diabetes age			
<40	2,685 (32.5%)	7,290 (8.9%)	P<0.0001
40-60	4,092 (49.6%)	28,860 (35.1%)	P<0.0001
>60	1,477 (17.9%)	46,025 (56%)	P<0.0001
# of ESRD	200 (2.4%)	600 (0.7%)	P<0.0001
ESRD age	56.5 ( $\pm$ 11.2)	64.1 ( $\pm$ 13.7)	P<0.0001
# of Death	1,482 (18%)	28,450 (34.6%)	P<0.0001
Death age	66.4 ( $\pm$ 14.4)	78.3 ( $\pm$ 11.1)	P<0.0001

## 6.2 Description of variables

The variables, codes and values are described in Table 6.2. Age at diabetes diagnosis is treated as a continuous variable.

Table 6.2. Description of variables.

Variables	Description	Codes/Values
DIABAGE	Age at diabetes diagnosis	Years (Continuous)
REGIND	Registered Indian Status	0=never, 1=ever
SEX	Sex	0=male, 1=female
TIME	Time since diabetes diagnosis	Years (Continuous)
EVENT	Types of events after diabetes development	0=censoring, 1=ESRD 2=death without ESRD

## 6.3 Non-parametric model

### 6.3.1 Cause-specific hazard rates

We can draw smoothed graphs of cause-specific hazard function for ESRD and death without ESRD from time of diabetes diagnosis. Figure 6.4 to Figure 6.6 present estimated smoothed hazard rate curves using the Epanechnikov kernel smoothed function. In these three graphs, bandwidth, abbreviated “ $bw$ ”, is the optimal estimation from SAS software analysis. Figure 6.4 presents the graph showing that the hazard rate for death was higher than that for ESRD, and we can also tell that hazard

rates increased over time (years since diabetes diagnosis) for both ESRD and death cases. Figure 6.5 presents crude smoothed hazard rates of ESRD over time between FN and OSK. The per-year risk of developing ESRD increased over time among FN and OSK. FN had a higher risk of ESRD than OSK. Figure 6.6 presents crude smoothed hazard rates of death over time between FN and OSK. The per-year risk of death increased over time among FN and OSK. For most of the study duration, the hazard rate of death was lower among FN than that of OSK.

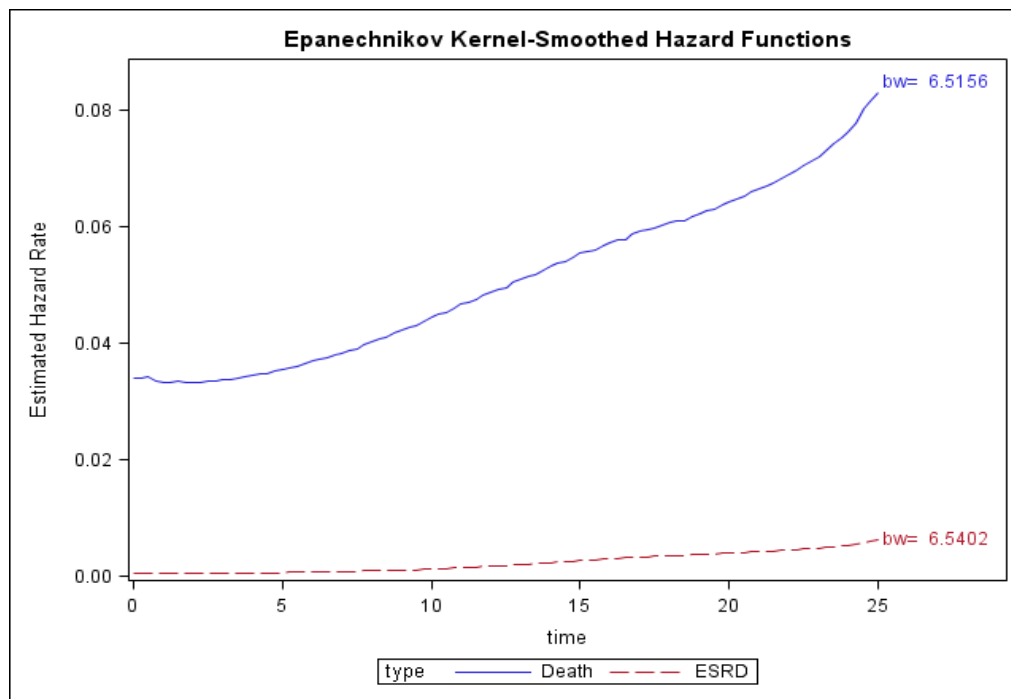


Figure 6.4 Overall hazard rates for ESRD and death cases.

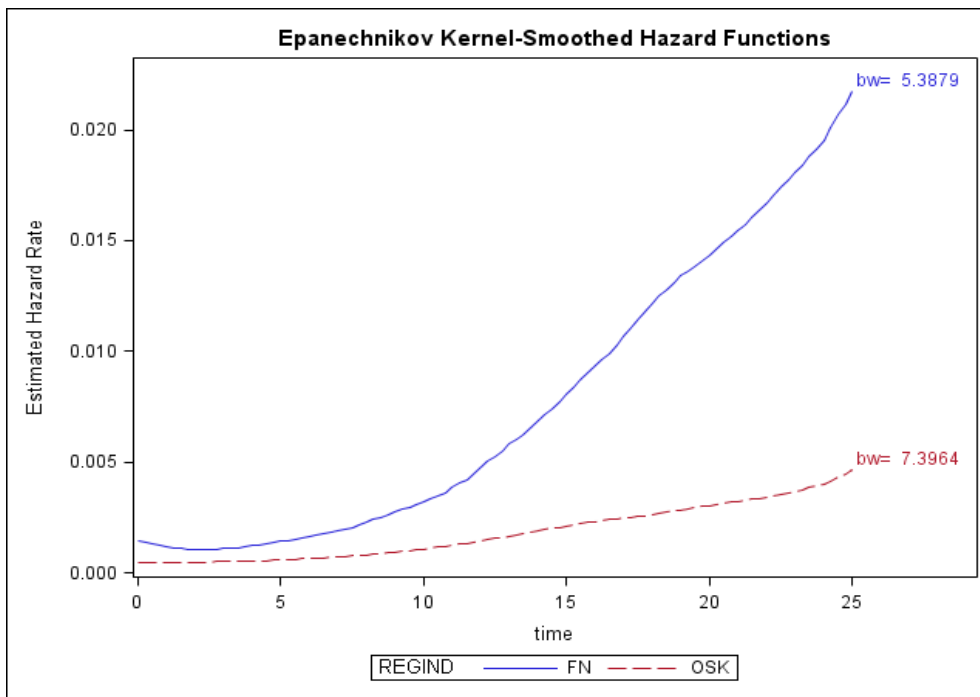


Figure 6.5 Hazard rates of ESRD by ethnicity.

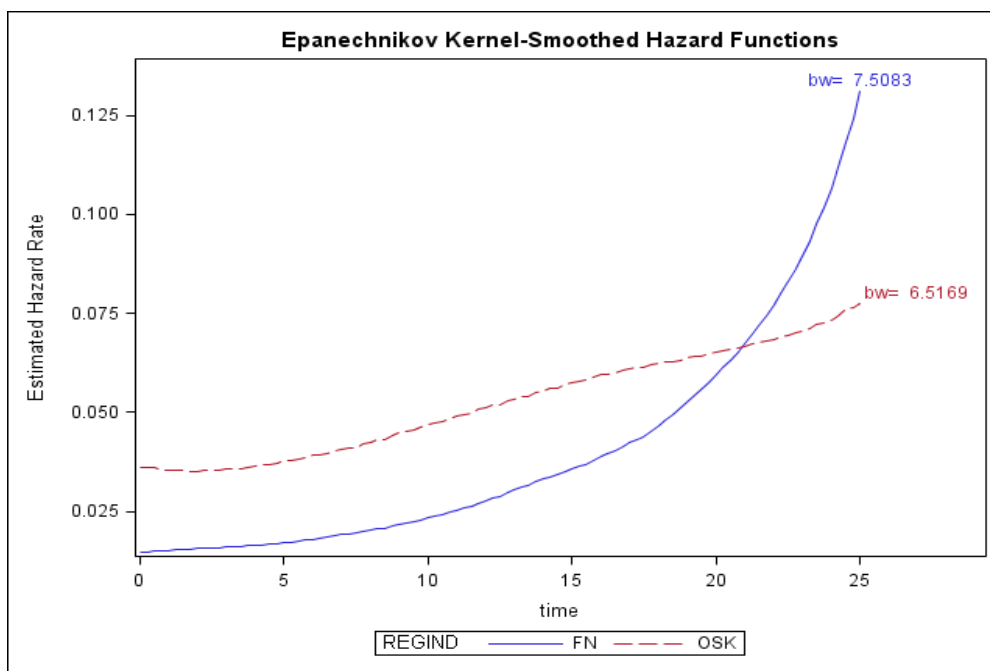


Figure 6.6 Hazard rates of death without ESRD by ethnicity.

### 6.3.2 Cumulative incidence function curves

Figure 6.7 and Figure 6.8 present the CIF curves for ESRD cases and death without



ESRD cases for FN and OSK, respectively, and the p-values for the differences between these two groups were both less than 0.0001. Overall, FN had a higher probability of ESRD and lower probability of death than OSK over time (years since diabetes diagnosis).

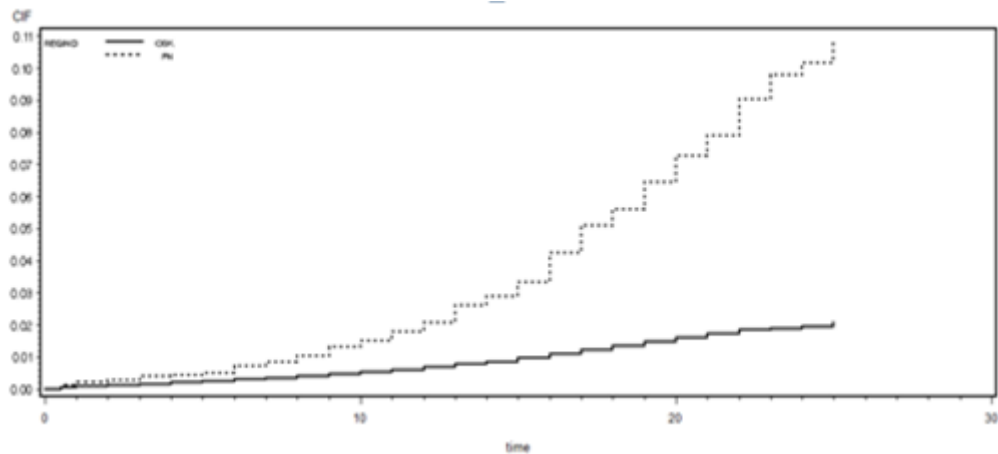


Figure 6.7 Cumulative incidence function (CIF) curve for ESRD by ethnicity.

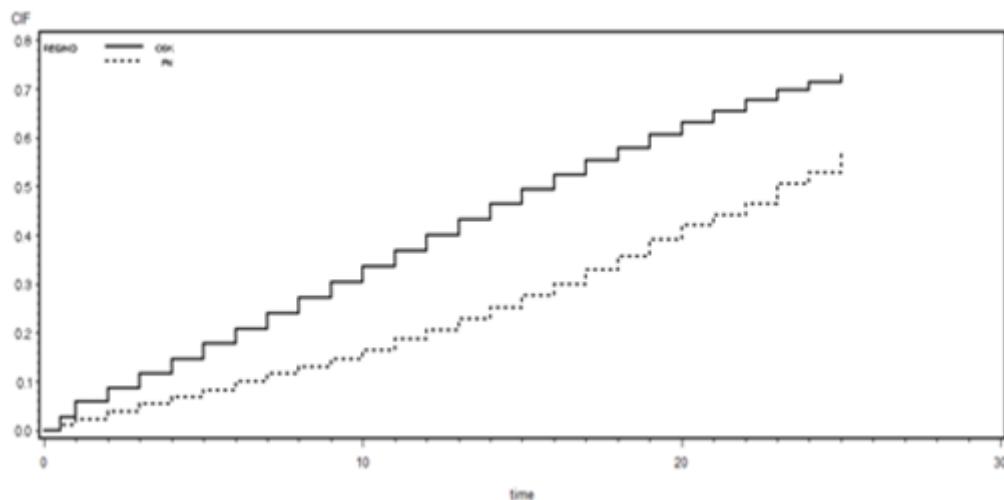


Figure 6.8 Cumulative incidence function (CIF) curve for death by ethnicity.

#### 6.4 Cox cause-specific model

Applying the univariable Cox cause-specific hazard model for ESRD and death with ESRD showed that sex, age at diagnosis of diabetes and ethnicity were significant

predictors (p-values<0.0001). For the ESRD outcome, the univariable model showed that the risk of developing ESRD was 3.19 times higher for FN compared to OSK (Table 6.3). For the death outcome, the univariable model showed that the risk of death for FN was 0.50 times that for OSK (Table 6.4).

Table 6.3. Univariable Cox cause-specific model for ESRD outcome.

Variables	$\beta$	Hazard Ratios	95% CI*	p-value
Male	0.44	1.55	(1.34,1.79)	<0.0001
First Nation	1.16	3.19	(2.72,3.75)	<0.0001
Diabetes age	-0.02	0.982	(0.977,0.986)	<0.0001

\* CI: Confidence Interval

Table 6.4. Univariable Cox cause-specific model for Death outcome.

Variables	$\beta$	Hazard Ratios	95% CI*	p-value
Male	0.20	1.22	(1.19, 1.24)	<0.0001
FN	-0.70	0.50	(0.47, 0.53)	<0.0001
Diabetes age	0.08	1.081	(1.080,1.082)	<0.0001

\* CI: Confidence Interval

After building multivariable models, we also tested the interactions between predictors in the model. The results show that for death without ESRD, there were significant interactions between ethnicity and age (P<0.0001) and between sex and age (P<0.0001) (Table 6.6), but no significant interaction for ESRD (Table 6.5). The final multivariable model showed that the risk of developing ESRD was 2.97 times higher

for FN compared to OSK (95% CI: 2.51-3.54,  $P < 0.0001$ ), adjusting for age at diabetes diagnosis and sex (Table 6.5). Most incident cases of diabetes occurred among FN in the age group of 40-49, while most new cases of diabetes occurred among OSK in the age group of 70+, leaving a nearly 25-year difference between FN and OSK in the age of diabetes diagnosis. The hazard ratios (HRs) of death without ESRD between FN males and OSK males with an age difference of 25 could be calculated as:

$$HR = \exp(1.19 - 25 \times 0.07 + 0.0147 \times 25) = 0.82$$

This HR is less than 1, which means that the risk of death is lower. Moreover, death HRs between FN females and OSK females with an age difference of 25 could be calculated as follows:

$$HR = \exp(1.19 - 25 \times 0.07 + 0.0147 \times 25 + 0.0063 \times 25) = 0.97$$

Table 6.5. Multivariable Cox cause-specific model for ESRD outcome.

Variables	$\beta$	Hazard Ratios	95% CI <sup>*</sup>	p-value
FN	1.09	2.97	(2.51,3.54)	<0.0001
Diabetes age	-0.01	0.99	(0.985,0.995)	<0.0001
Male	0.49	1.64	(1.42,1.89)	<0.0001

\* CI: Confidence Interval

Table 6.6. Multivariable Cox cause-specific model for death outcome.

Variables	$\beta$	Hazard Ratios	95% CI <sup>*</sup>	p-value
FN	1.19	3.30	(2.63,4.14)	<0.0001
Diabetes age	0.07	1.083	(1.081,1.083)	<0.0001
Male	0.82	2.27	(1.98,2.61)	<0.0001

Diabetes age*FN	-0.0147	0.985	(0.982,0.989)	<0.0001
Diabetes age*Male	-0.0063	0.994	(0.992, 0.996)	<0.0001

\*CI: Confidence Interval

## 6.5 Parametric proportional hazards model

Table 6.7 present AIC values (discussed in Section 5.2.4.3) for two parametric PH models for ESRD and

Table 6.8 present AIC for two parametric PH models for death without ESRD respectively: the Weibull PH model and the exponential PH model. Model fitting is better if AIC values are lower. Therefore, the Weibull PH model was selected for both ESRD and death without ESRD.

Table 6.7. AIC in Parametric PH models for ESRD.

Parametric model	AIC
Exponential	8,975
Weibull	8,704

Table 6.8. AIC in Parametric PH for death without ESRD.

Parametric model	AIC
Exponential distribution	128,276

Weibull distribution	123,452
----------------------	---------

Table 6.9 presents the results from the Weibull PH model for ESRD; we found similar hazard ratios ( $\exp(1.07)=2.89$ ), close to that from the Cox cause-specific hazard model between FN and OSK after adjusting for age and sex. For ESRD cases, the estimations of scale and intercept,  $\sigma$  and  $\mu$ , were 0.60 and 4.66, respectively. The hazard function for the  $i$ th person,  $h_i(t)$ , from the Weibull PH model can be calculated from equation (5.7) to (5.10):

$$h_i(t) = 0.0004 \times 1.66t^{0.66} \exp(-0.01DIABAGE + 1.07REGIND - 0.48SEX)$$

Similarly, the hazard rate function for death without ESRD using the Weibull PH model results from \*CI: Confidence Interval

Table 6.10 is:

$$h_i(t) = 0.0001 \times 1.41t^{0.41} \exp(0.08DIABAGE + 1.17REGIND - 0.79SEX - 0.014REGIND * DIABAGE + 0.006SEX * DIABAGE)$$

Hazard rates can be estimated with covariate information. For example, the hazard rate of ESRD for 45-year-old FN females with diabetes is:

$$h(t) = 0.0004 * 1.66t^{0.66} \exp(-0.01 * 45 + 1.07 - 0.48) = 0.0008t^{0.66}$$

with the hazard rate increasing over time. Moreover, the hazard rate of death without ESRD for 70-year-old FN females with diabetes is:

$$h(t) = 0.0001 * 1.41t^{0.41} \exp(0.08 * 70 + 1.17 * 1 - 0.79 * 1 - 0.014 * 70 + 0.006 * 70) = 0.039 t^{0.41}$$

with the hazard rate increasing over time.

Table 6.9. Weibull PH model for ESRD.

Variable	$\beta$	Hazard Ratios	95% CI*	p-value
Intercept	4.66			<0.0001
Diabetes age	-0.01	0.99	(0.982, 0.992)	<0.0001
FN	1.07	2.91	(2.51, 3.36)	<0.0001
Male	0.48	1.62	(1.36, 1.94)	<0.0001
Scale	0.60			
Shape	1.66			

\*CI: Confidence Interval

Table 6.10. Weibull PH model for death without ESRD.

Variable	$\beta$	HR	95% CI*	p-value
Intercept	6.40			<0.0001
Diabetes age	0.08	1.08	(1.079, 1.082)	<0.0001
FN	1.17	3.22	(2.79, 3.71)	<0.0001
Male	0.79	2.20	(1.75, 2.76)	<0.0001
FN*Diabetes age	-0.014	0.986	(0.98, 0.99)	<0.0001
Male*Diabetes age	-0.006	0.994	(0.992, 0.996)	<0.0001

---

Scale	0.71
Shape	1.40

---

\*CI: Confidence Interval

## 6.6 Piece-wise exponential proportional hazards model

The piece-wise exponential distributions can be used to test the effect of covariates and also to estimate the hazard rates of events over time given the effects of other covariates. After fitting data to piece-wise exponential model, we found hazard rates increased overall and within each time interval, the rate was constant. Furthermore, in section 6.5, estimated hazard rates from Weibull PH model increased over time and in section 5.3, one assumption for SDM requires hazard rate being constant. The piece-wise exponential PH model incorporates the characteristics of the Weibull PH model prediction and the assumptions for SDM.

We modeled the survival data with 3-year and 5-year intervals for ESRD (Table 6.11 and Table 6.12 respectively). For ESRD cases, if we divided time into 5-year intervals, we find that the hazard rate increased with time monotonically. If we divided time into 3-year intervals, the hazard rate mostly increased over time monotonically, but there was a slight decrease in the 8<sup>th</sup> time interval. For death cases, Table 6.14 shows that with 5-year intervals, the hazard rate increased monotonically; for 3-year intervals (Table 6.13), the hazard rate usually increased monotonically, except that the hazard rate decreased slightly during the 2<sup>nd</sup> interval and then increased throughout the following intervals. The hazard ratios between FN and OSK were close to the results

from the Cox cause-specific models for both ESRD and death without ESRD. For example, the hazard ratio of ESRD between FN and OSK was  $\exp(1.09)=2.97$  using 3-year intervals or 5-year intervals, which was similar to the Cox cause-specific model results.

The estimated hazard rate function of ESRD for the  $i$ th person,  $h_i(t)$ , from the piece-wise exponential PH model (3-year intervals) can be calculated from equations (5.15-5.16),  $j_1, j_2, \dots, j_8$  are indicators for intervals. When  $j_1$  is equal to one and  $j_2, j_2, \dots, j_8$  are equal to zero, we could get hazard rate estimations for the second interval, the third year to the sixth year.

$$h_i(t) = \exp(-4.55 - 2.17j_1 - 2.21j_2 - 1.93j_3 - 1.5j_4 - 1.07j_5 - 0.48j_6 - 0.34j_7 - 0.45j_8) * \exp(-0.01DIABAGE + 1.09REGIND - 0.49SEX)$$

Similarly, the estimated hazard rate of death without ESRD from the piece-wise exponential PH model (3-year intervals) for the  $i$ th person is:

$$h_i(t) = \exp(-6.71 - 1.75j_1 - 1.76j_2 - 1.47j_3 - 1.16j_4 - 0.86j_5 - 0.59j_6 - 0.30j_7 - 0.06j_8) * \exp(0.08DIABAGE + 1.20REGIND - 0.82SEX - 0.015REGIND * DIABAGE + 0.01SEX * DIABAGE)$$

Given the above two formulas, the hazard rates of ESRD for 45-year-old FN males with diabetes in the first and third time intervals (0-3rd, 6th-9th year) were 0.0023 and 0.0029 respectively:

$$h(t) = \exp(-4.55 - 2.17 * 1) * \exp(-0.01 * 45 + 1.09 * 1 - 0.49 * 0) = 0.0023$$

$$h(t) = \exp(-4.55 - 1.93 * 1) * \exp(-0.01 * 45 + 1.09 * 1 - 0.49 * 0) = 0.0029$$

The hazard rates of death without ESRD for 45-year-old FN males with diabetes in the first and third time intervals (0-3rd, 6th-9th year) were 0.013 and 0.017 respectively:

$$h(t) = \exp(-6.71 - 1.75 * 1) * \exp(0.08 * 45 + 1.2 * 1 - 0.82 * 0 - 0.015 * 1 * 45) = 0.013$$



$$h(t)=\exp(-6.71-1.47*1)*\exp(0.08*45+1.2*1-0.82*0-0.015*1*45)=0.017$$

Table 6.11. Piece-wise exponential model for ESRD outcome with 3-year interval.

Variable	$\beta$	95% CI*	P-value
Intercept	4.55	(3.55, 5.56)	<.0001
Male	0.49	(0.35, 0.64)	<.0001
FN	1.09	(0.92, 1.26)	<.0001
Diabetes age	-0.01	(-0.016,-0.006)	<.0001
j1 [3, 6)	2.17	(1.17, 3.16)	<.0001
j2 [6, 9)	2.21	(1.21, 3.21)	<.0001
j3 [9, 12)	1.93	(0.92, 2.93)	0.0002
j4 [12, 15)	1.50	(0.50, 2.50)	0.0033
j5 [15, 18)	1.07	(0.072, 2.07)	0.0356
j6 [18, 21)	0.48	(-0.51, 1.48)	0.3435
j7 [21, 24)	0.34	(-0.67, 1.34)	0.5106
j8 [24, 25)	0.45	(-0.60, 1.50)	0.4037

\*CI: Confidence Interval

Table 6.12. Piece-wise exponential model for ESRD outcome with 5-year interval.

Variable	$\beta$	95% CI*	P-value
Intercept	4.93	(4.56, 5.31)	<.0001

Male	0.49	(0.35, 0.64)	<.0001
FN	1.09	(0.92, 1.26)	<.0001
Diabetes age	-0.01	(-0.016, -0.006)	<.0001
j1 [5, 10)	1.83	(1.52, 2.15)	<.0001
j2 [10, 15)	1.53	(1.21, 1.85)	<.0001
j3 [15, 20)	0.81	(0.49, 1.12)	<.0001
j4 [20, 25)	0.05	(-0.26, 0.36)	0.7585

\*CI: Confidence Interval

Table 6.13. Piece-wise exponential model for death outcome with 3-year interval.

Variable	$\beta$	95% CI*	P-value
Intercept	6.71	(6.38, 7.05)	<.0001
Male	0.82	(0.68, 0.96)	<.0001
FN	1.20	(0.97, 1.43)	<.0001
Diabetes age	0.08	(0.078, 0.081)	<.0001
Male*Diabetes age	-0.006	(0.0043, 0.0082)	<.0001
FN*Diabetes age	-0.015	(0.011, 0.019)	<.0001
j1 [3, 6)	1.75	(1.43, 2.07)	<.0001
j2 [6, 9)	1.76	(1.44, 2.09)	<.0001
j3 [9, 12)	1.47	(1.15, 1.79)	<.0001
j4 [12, 15)	1.16	(0.84, 1.48)	<.0001

j5 [15, 18)	0.86	(0.53, 1.18)	<.0001
j6 [18, 21)	0.59	(0.26, 0.91)	0.0004
j7 [21, 24)	0.30	(-0.025, 0.63)	0.0702
j8 [24, 25)	0.06	(-0.27, 0.40)	0.7152

\*CI: Confidence Interval

Table 6.14. Piece-wise exponential model for death outcome with 5-year interval.

Variable	$\beta$	95% CI*	P-value
Intercept	6.83	(6.72, 6.94)	<.0001
Male	0.82	(0.68, 0.96)	<.0001
FN	1.19	(0.96, 1.42)	<.0001
Diabetes age	0.08	(0.0777, 0.0804)	<.0001
Male*Diabetes age	-0.01	(-0.0082, -0.0043)	<.0001
FN*Diabetes age	-0.015	(- 0.018, -0.011)	<.0001
j1 [5, 10)	1.62	(1.55, 1.70)	<.0001
j2 [10, 15)	1.34	(1.26, 1.41)	<.0001
j3 [15, 20)	0.84	(0.76, 0.91)	<.0001
j4 [20, 25)	0.37	(0.29, 0.45)	<.0001

\*CI: Confidence Interval

## 6.7 Computational Modeling

### 6.7.1 System Dynamics modeling

FigureTable 6.9 presents one System Dynamics model with a 5-year interval, leaving 5 stocks that are labeled “diabetes 1”, “diabetes 2”, “diabetes 3”, “diabetes 4” and “diabetes 5”. Another two stocks presented in the model are called “ESRD” and “Death without ESRD”. Diabetes progression is represented by flows such as “diabetes progression 2”, “diabetes progression 3”, “diabetes progression 4” and “diabetes progression 5”. Diabetic people may develop ESRD or die without developing ESRD. Flows for ESRD development are shown in the model by “ESRD incidence 1” through “ESRD incidence 5”, and similarly, flows from diabetes to death without ESRD are indicated as “Death at diabetes interval 1” through “Death at diabetes interval 5”.

The values of flows are ultimately determined by stock values and constant rate values. For example, in this model, ESRD incidence 1 flow is determined as the stock value of “diabetes 1” times the constant rate “ESRD rate 1”. In this model, we used a subgroup method to stratify the model by FN males, FN females, OSK males, and OSK females. The rates of ESRD and death were both subgrouped by ethnicity and sex. In our analysis, age at diabetes diagnosis was treated as a continuous variable, so we also used subgroups to present age at diabetes diagnosis categories from age 20 to age 107.

The structure of the model also included competing risk effects of ESRD and death among the diabetic population. During each year of the study period, there were incident cases of diabetes coming into the model by entering the stock of “diabetes 1”.

Once people entered the “diabetes 1” stock, there were three stocks to which they could transfer: “ESRD”, “Death without ESRD” and “diabetes 2” (the last of which represented the second stock of the diabetic population). The mean time for people to move from stock “diabetes 1” to “diabetes 2” was 5 years. Again, for people who entered the diabetes 2 stock, they could go in three directions: ESRD, death without ESRD and “diabetes 3”. The same applied for people in the stock of “diabetes 3” or “diabetes 4”. Eventually, if people had diabetes for over 20 years, which means that they would enter the stock of “diabetes 5”, there were only 2 directions for them to go: “ESRD” and “Death without ESRD”. The 3-year interval model was similar to this 5-year interval model but required 4 more stocks for diabetes, flows between stocks, ESRD rates from diabetes to ESRD, and death rates from diabetes to death.

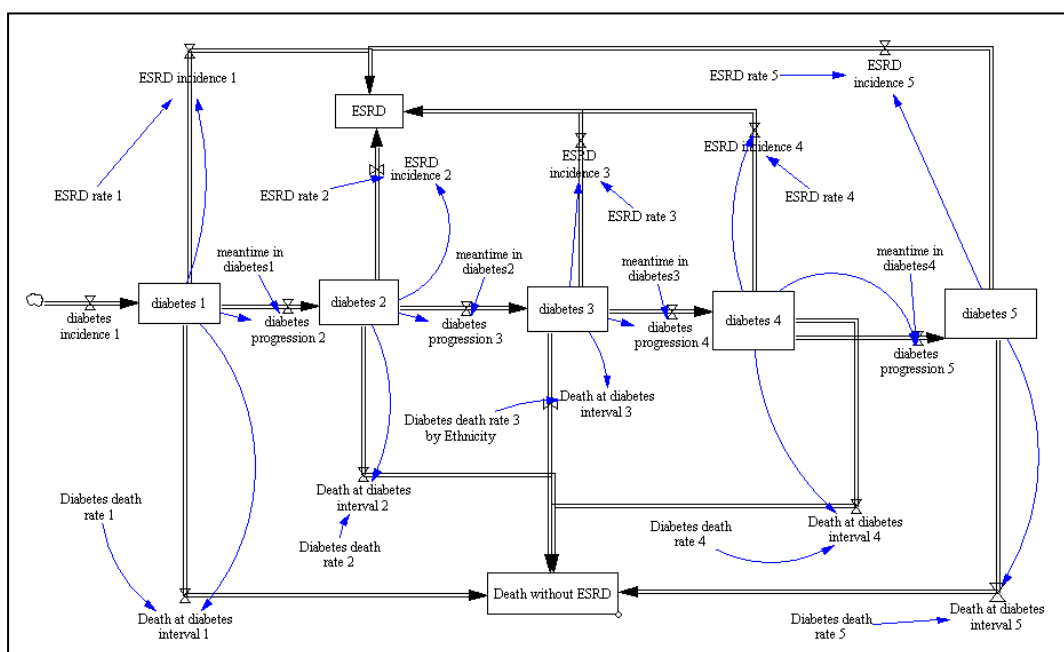


Figure 6.9 System Dynamics Model with a 5-year interval.

To check the goodness of fit of our dynamic modeling predictions, we compared

the model prediction with historical data and drew graphs in EXCEL. Figure 6.10 shows the comparison between historical data and model data for the ESRD event using 3-year piece-wise exponential estimations from survival analysis. In this figure, after around year 1996, there was a discrepancy between historical data and model data. A similar goodness of fit pattern is shown in Figure 6.11, which shows the comparison between historical data and model data for ESRD using 5-year piece-wise exponential PH survival analysis, and there was a certain discrepancy during the entire study period. Figure 6.12 and Figure 6.13 present a comparison between historical data and model data for death without ESRD. The 5-year interval model overestimated death incidence cases for most of the study duration, while 3-year interval estimations from the model showed that after year 1992 since diabetes diagnosis, the incident cases of deaths were underestimated by the model.

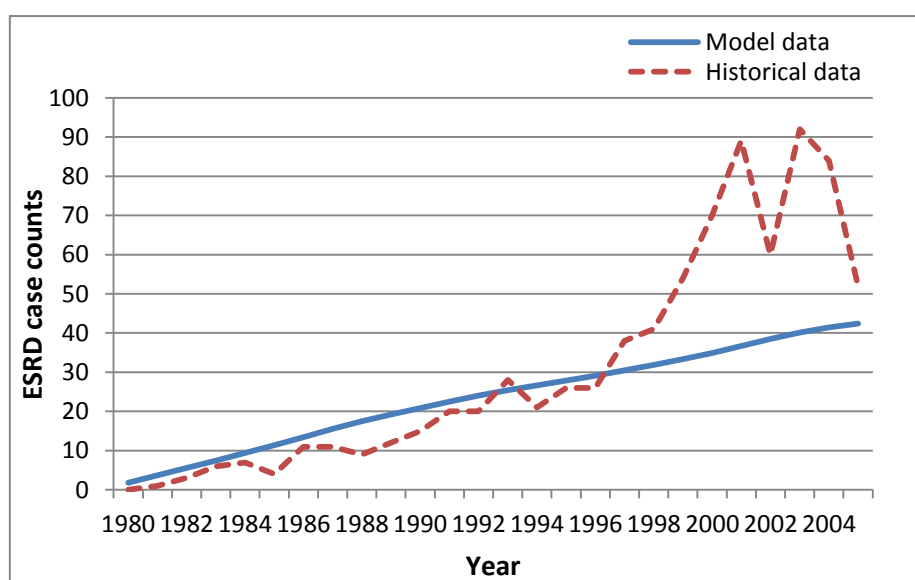


Figure 6.10 Comparison of historical data with model-estimated data with a 3-year interval analysis of a piece-wise exponential model for ESRD.

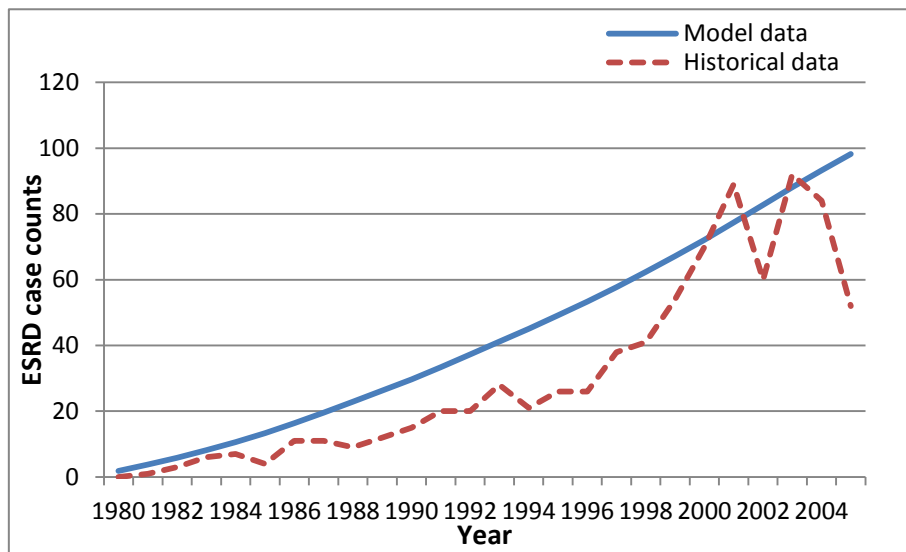


Figure 6.11 Comparison of historical data with model-estimated data with a 5-year interval analysis of a piece-wise exponential model for ESRD.

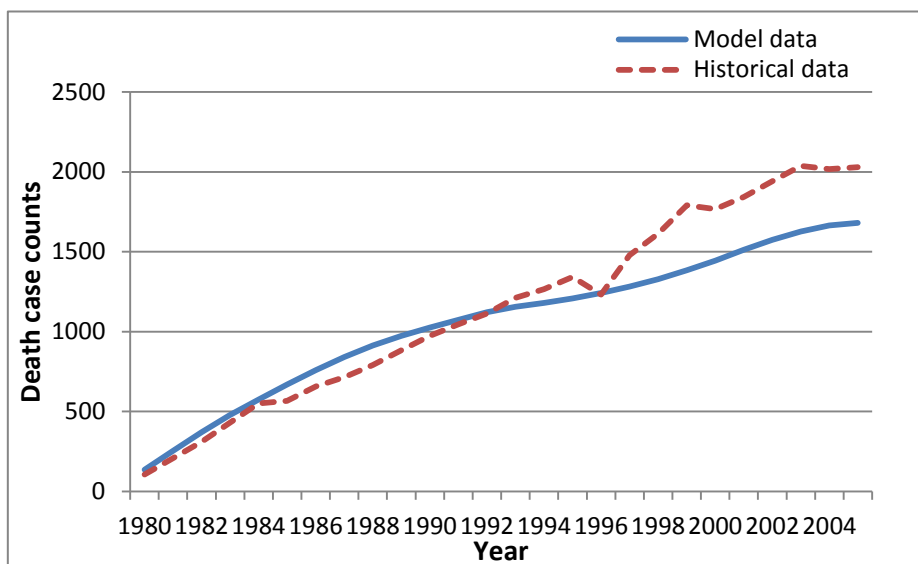


Figure 6.12 Comparison historical data with model estimated data with 3-year interval analysis of piece wise exponential for death without ESRD.

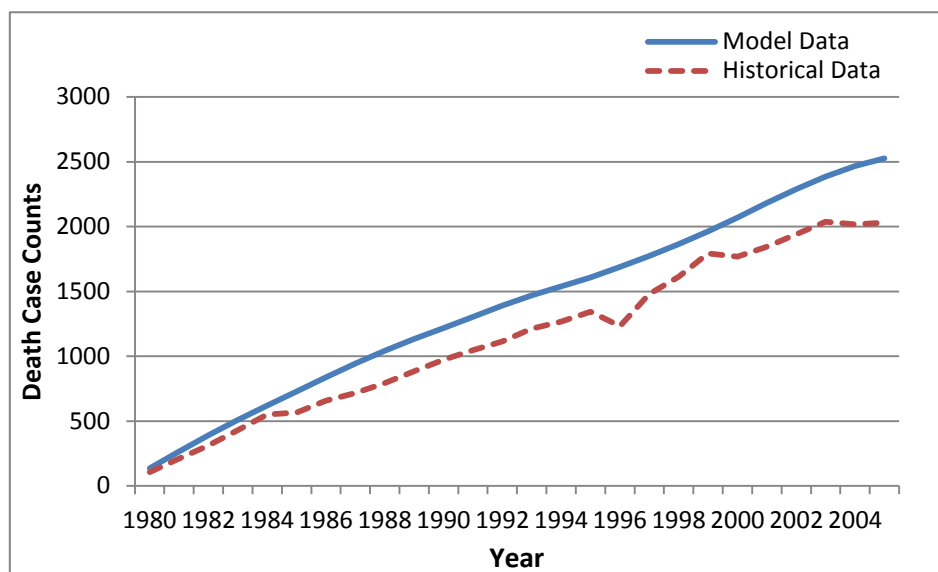


Figure 6.13 Comparison of historical data with model-estimated data with a 5-year interval analysis of a piece-wise exponential model for death without ESRD.



### 6.7.2 Agent-based modeling

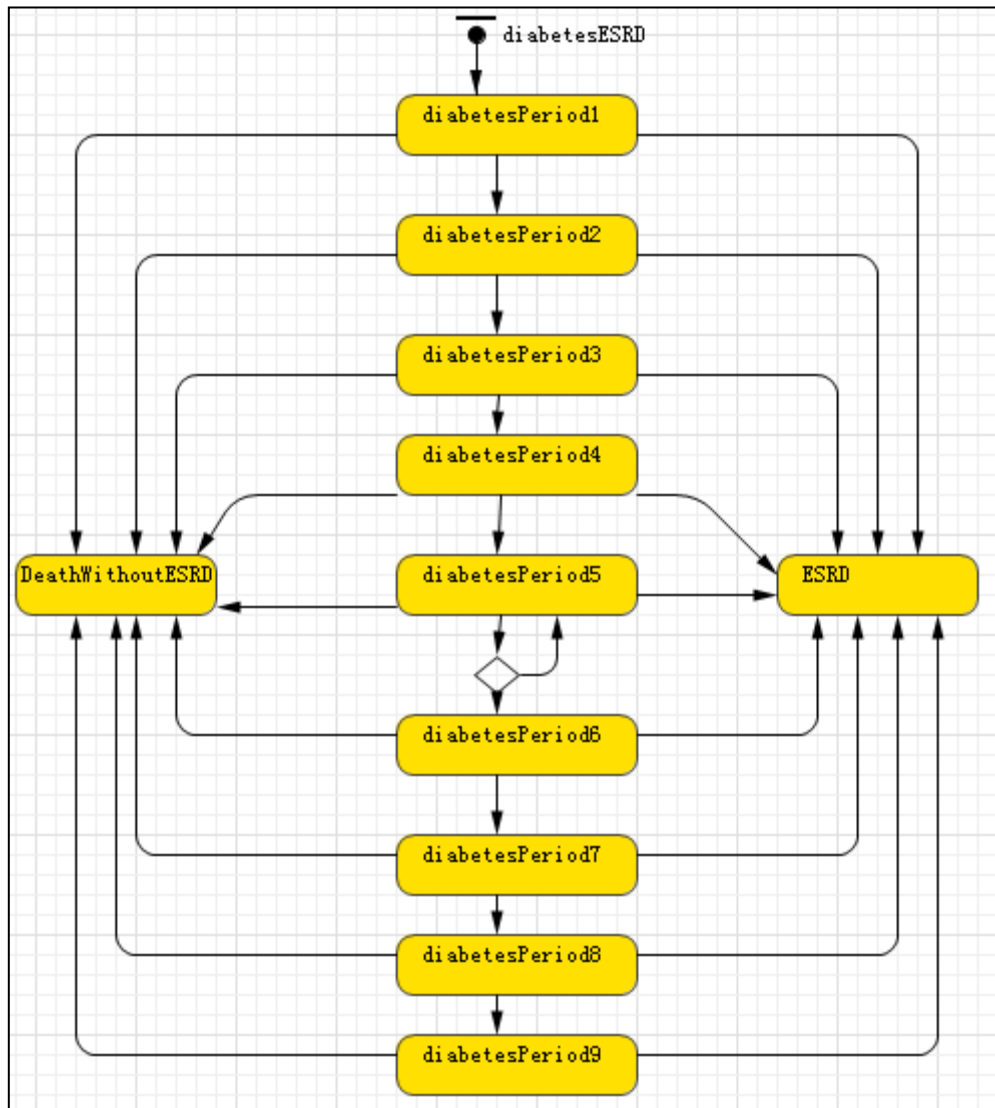


Figure 6.14 State chart of the agent-based model for person flow.

Figure 6.14 presents one state chart of competing risks effects of ESRD and death without ESRD. There are 9 states for diabetes. For each diabetes state, there are two other states connected with it, that is, “ESRD” and “DeathWithoutESRD”. Links between states are transitions. ESRD and death here are also competing risk events, and those people who transition to the “ESRD” state cannot enter “DeathWithoutESRD”

because death represents death without ESRD. Similarly, a person who transitions from diabetes to “DeathWithoutESRD” will not enter the “ESRD” state. Transitions from diabetes state to “ESRD” or from diabetes state to “DeathwithESRD” are the estimated hazard rates. Transitions between diabetes states are timeout such as 3 years or 5 years. There is a branch between state 5 and state 6, and this function is used to provide flexibility to support both 3-year and 5-year statecharts. Specifically, if the time for people to stay in one statechart is 5 years, the outward transition goes back to “diabetesPeriod5” (i.e., the person remains in “diabetesPeriod5”); otherwise, the diabetic person moves on to “diabetesPeriod6” and then progresses to later states (up to “diabetesPeriod9”).

Figure 6.15 to Figure 6.18 compared historical data and model predictions from one simulation experiment, which matched well for both ESRD and death cases when we used ABM with either a 3-year or a 5-year interval. One point needs to be stated here. The historical ESRD case curve dropped near the end of the study. One possible reason for this is that the ESRD definition needs a 3-month duration for dialysis. In the data we had, the number of ESRD cases was underestimated for the year 2005.

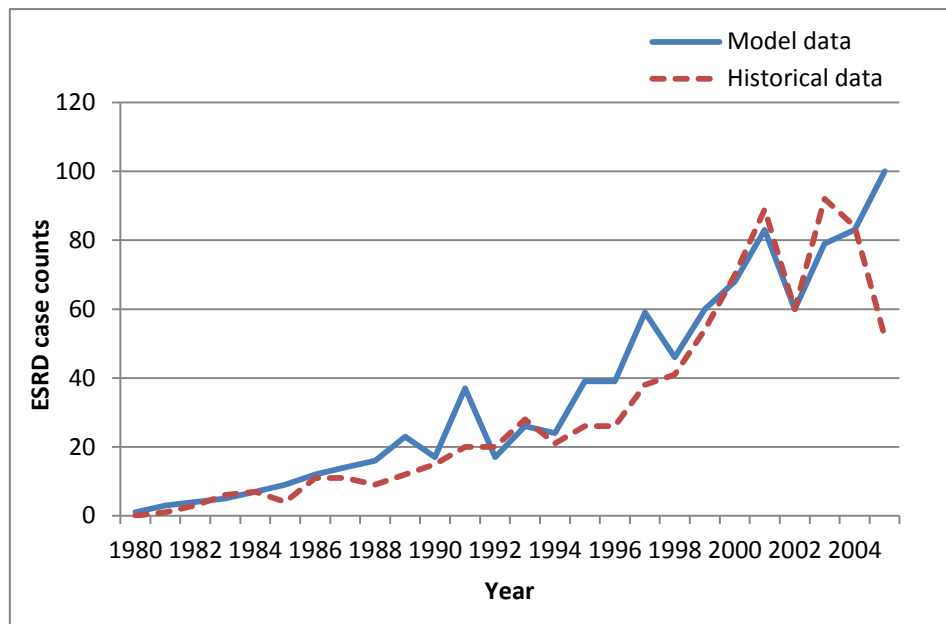


Figure 6.15 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for ESRD.

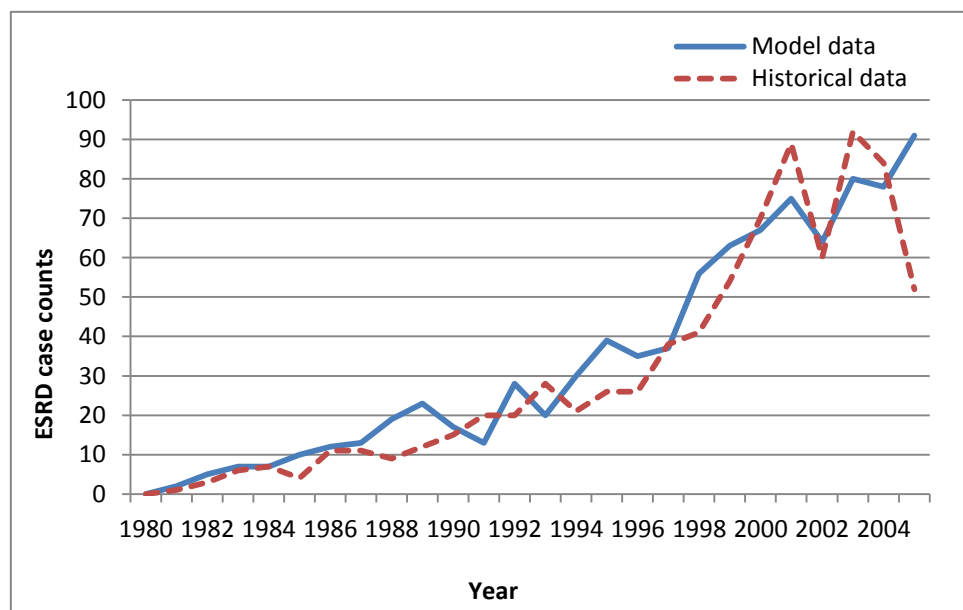


Figure 6.16 Comparison between historical data and model-estimated data with a 5-year interval analysis of a piece-wise exponential model for ESRD.

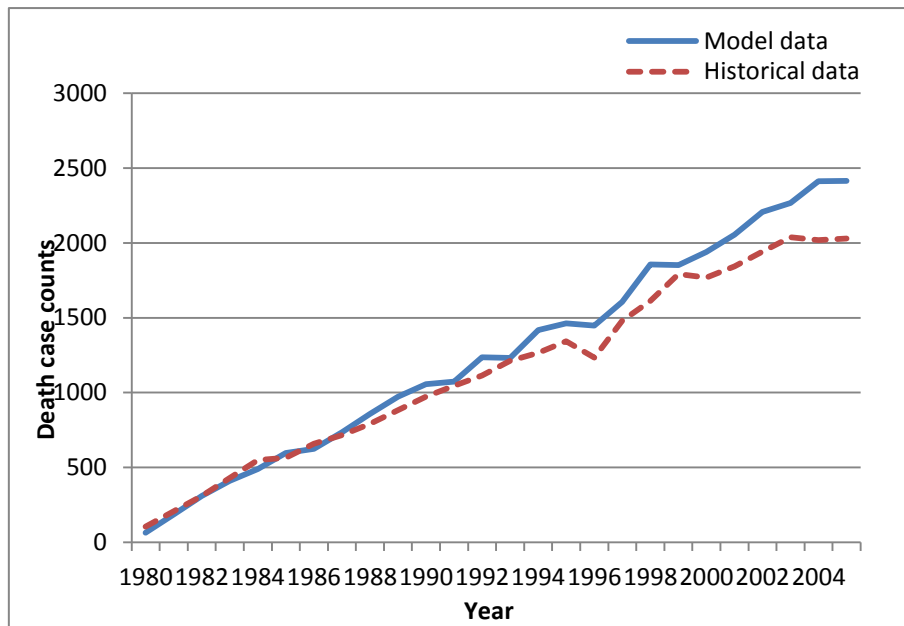


Figure 6.17 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for death without ESRD.

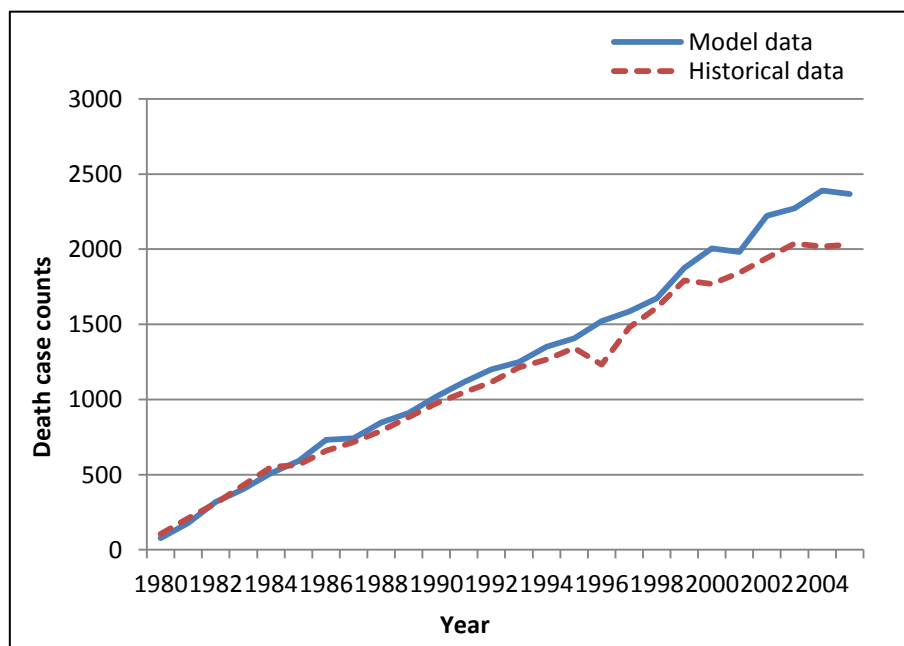


Figure 6.18 Comparison between historical data and model-estimated data with a 5-year interval analysis of a piece-wise exponential model for death without ESRD.

To further analyze the data, we performed a subgroup analysis. Figure 6.19 to Figure 6.22 present comparisons of ESRD between the historical data and model predictions for FN males, FN females, OSK males, and OSK females, respectively. The results show that all of the data matched well. Similar results were found for death in these four groups (Figure 6.23 to Figure 6.26).

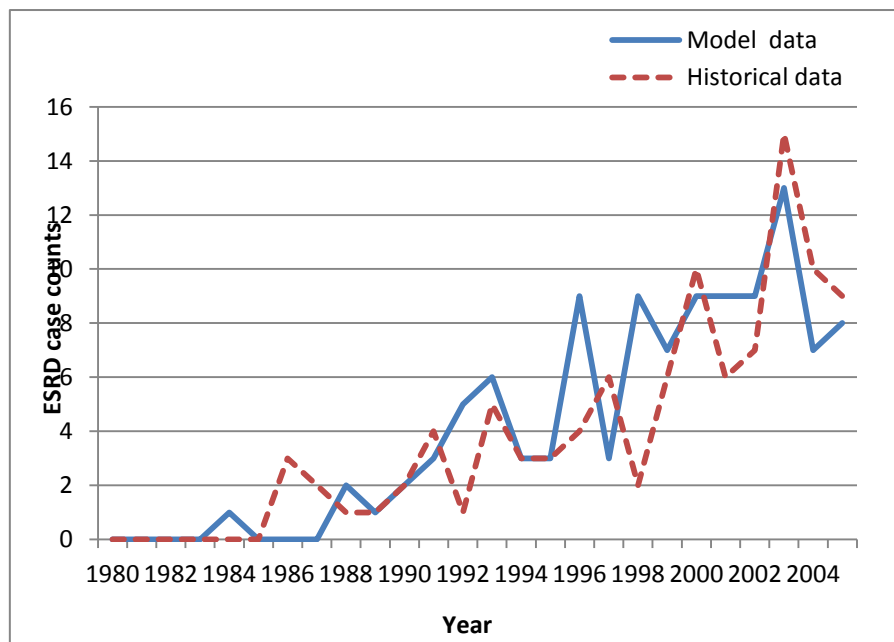


Figure 6.19 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for ESRD in FN males.

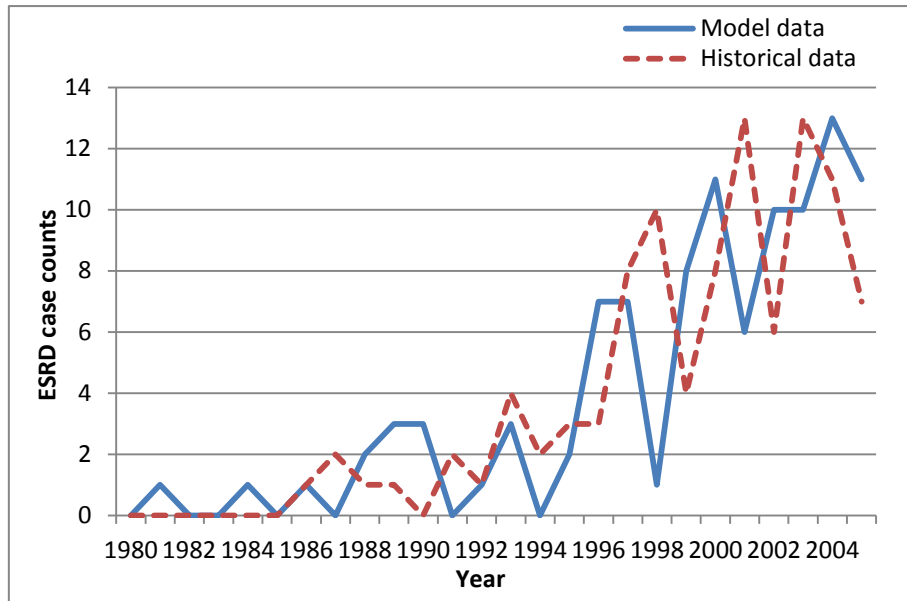


Figure 6.20 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for ESRD in FN females.

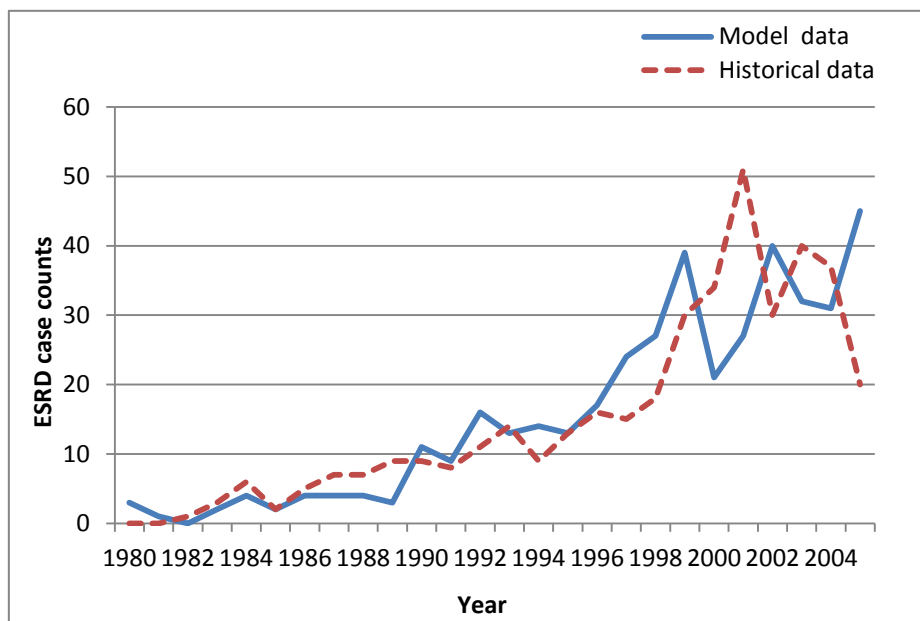


Figure 6.21 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for ESRD in OSK males.

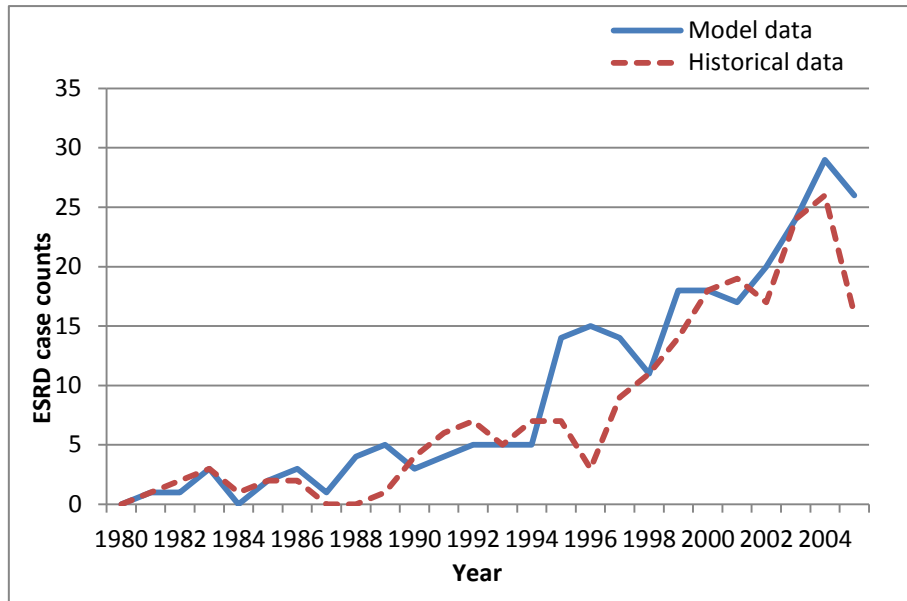


Figure 6.22 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for ESRD in OSK females.

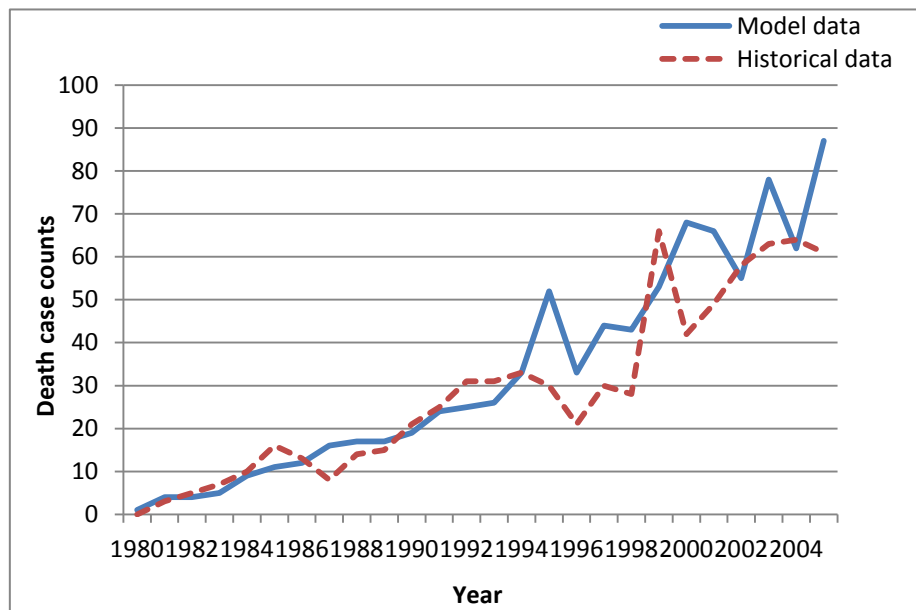


Figure 6.23 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for death without ESRD in FN males.

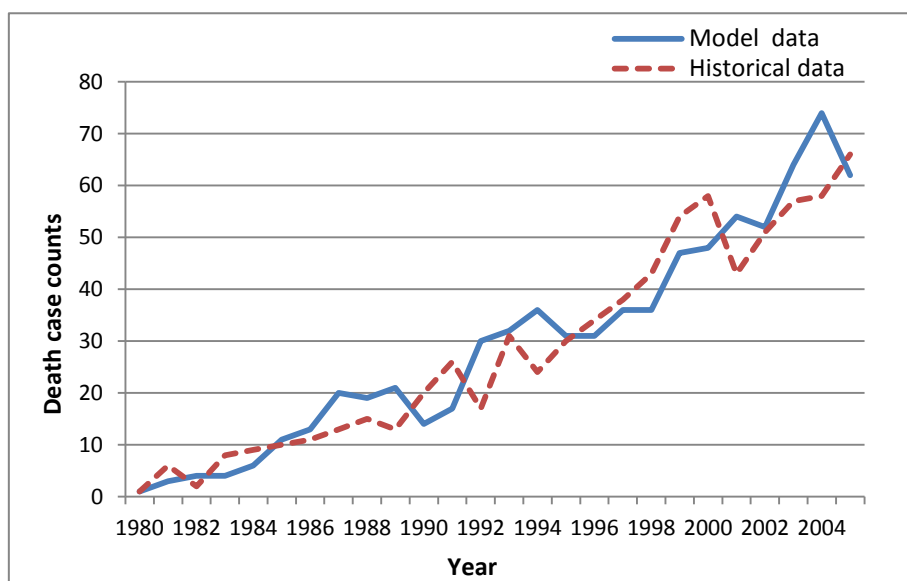


Figure 6.24 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for death without ESRD in FN females.

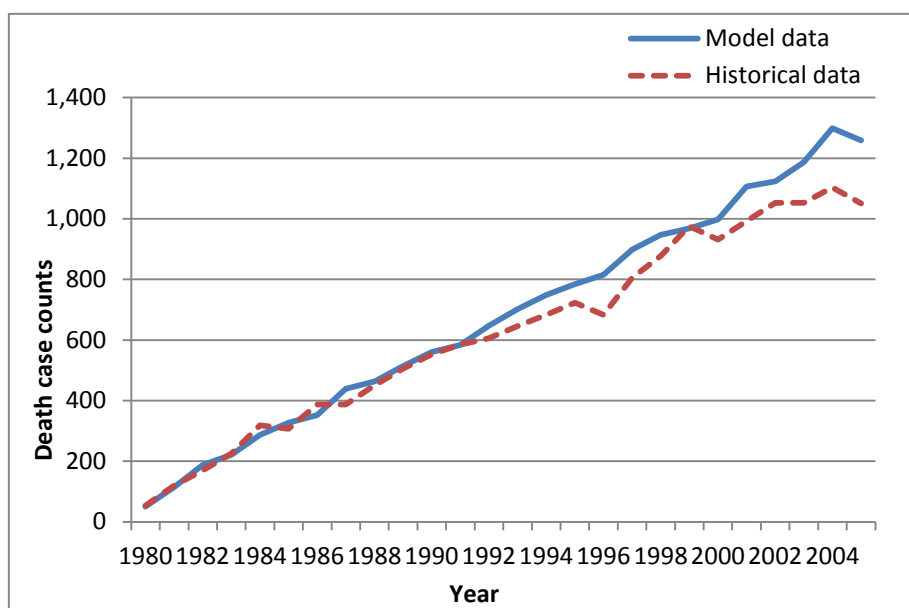


Figure 6.25 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential model for death without ESRD in OSK males.



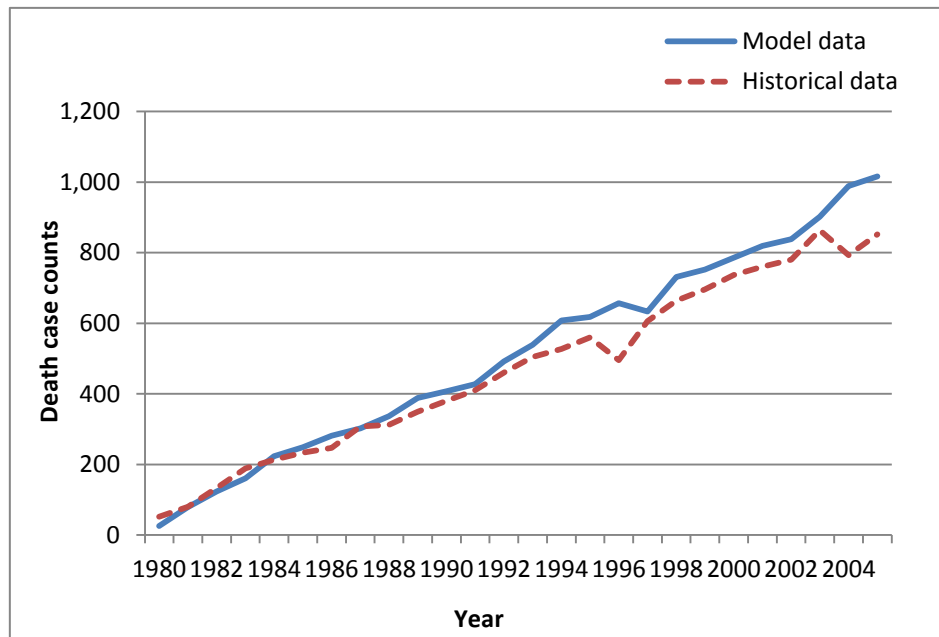


Figure 6.26 Comparison between historical data and model-estimated data with a 3-year interval analysis of a piece-wise exponential for death without ESRD in OSK females.

The above graphs are based on a single simulation realization of the agent-based model. We also performed Monte Carlo analysis based on the model with 100 simulation realizations. Figure 6.27 and Figure 6.28 present comparisons between historical and model-predicted data for overall ESRD and death without ESRD incident cases using a 3-year interval piece-wise exponential PH model. The gray area represents agent-based dynamic model predication data (given 100 realizations), and the red line presents historical data. The darkest color indicates 2.5% simulation results and the 97.5% simulation results are marked with lightest grey area. The figures show good matches between these model data and historical data. Figure 6.29 to Figure 6.36 represent similar comparisons for subgroups of FN males, FN females, OSK males, and OSK females for both events, also using 3-year estimations, and there were good

matches between historical data and model-predicted data.

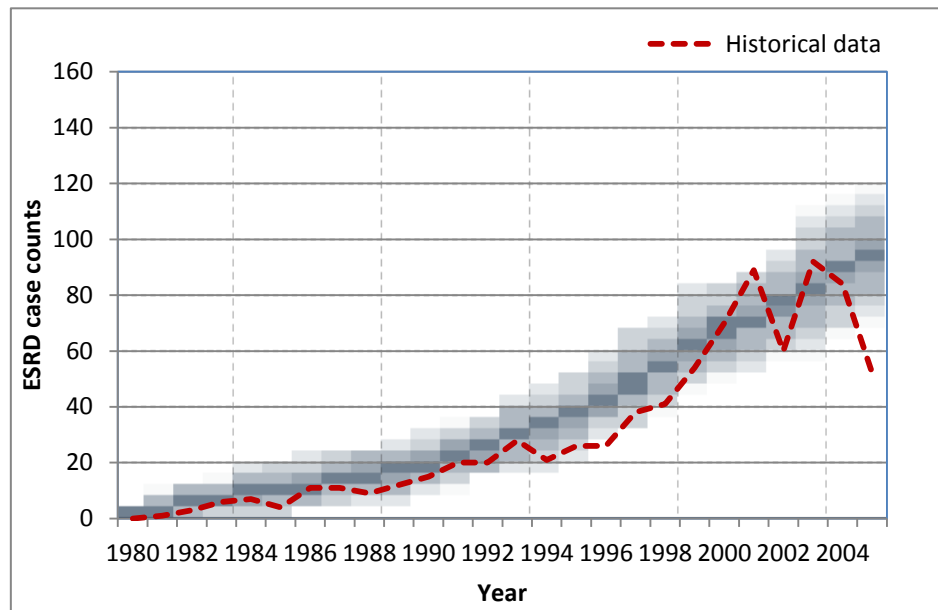


Figure 6.27 Comparison between incident historical data and model-estimated data for ESRD from a Monte Carlo model.

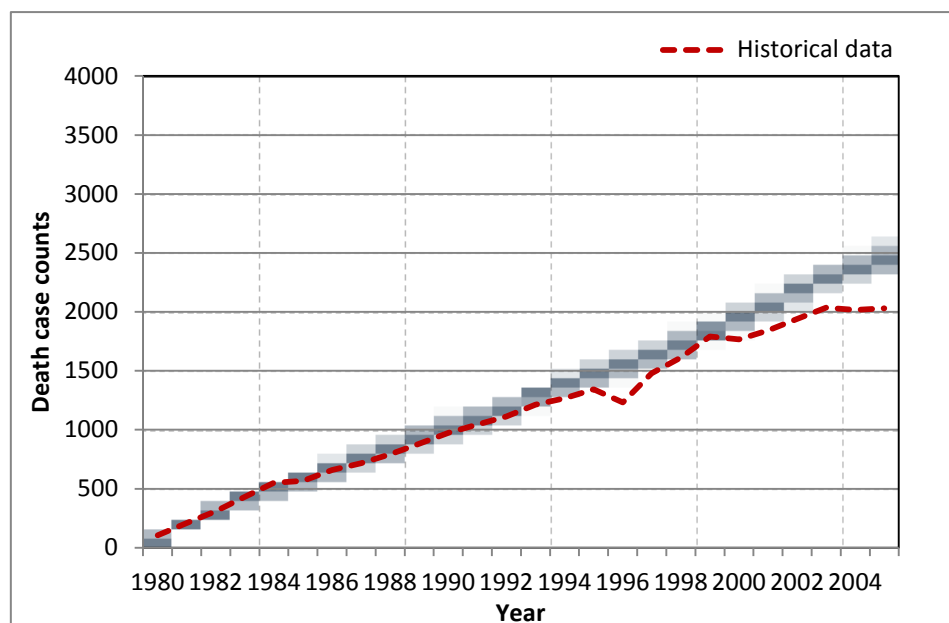


Figure 6.28 Comparison between incident historical data and model-estimated data for death without ESRD from a Monte Carlo model.

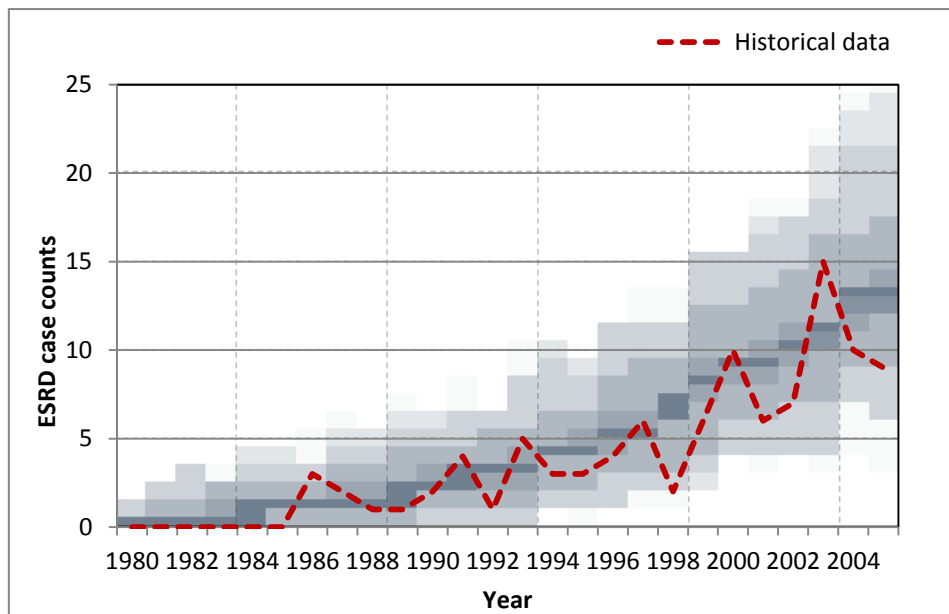


Figure 6.29 Comparison between incident historical data and model-estimated data for ESRD in FN males from a Monte Carlo model.

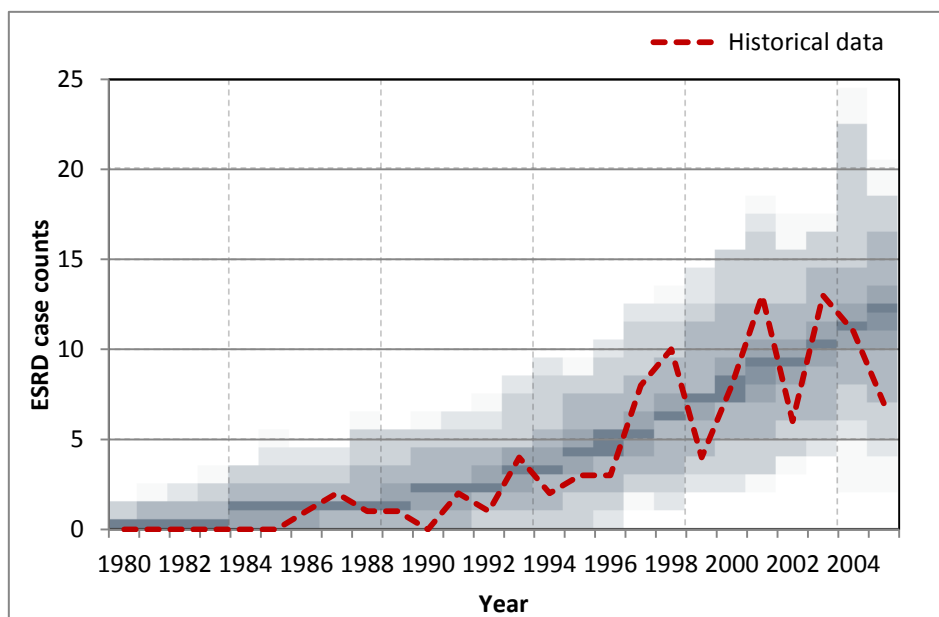


Figure 6.30 Comparison between incident historical data and model-estimated data for ESRD in FN females from a Monte Carlo model.

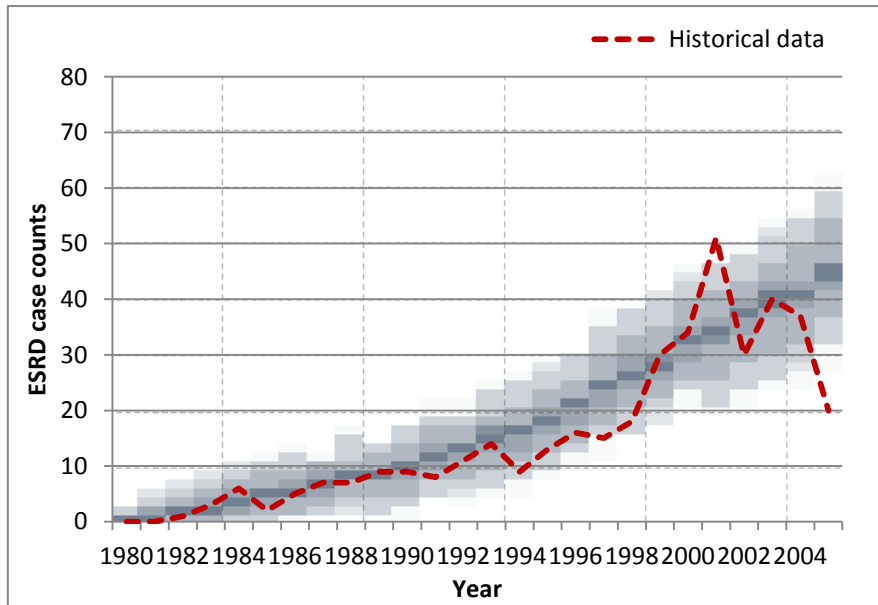


Figure 6.31 Comparison between incident historical data and model-estimated data for ESRD in OSK males from a Monte Carlo model.

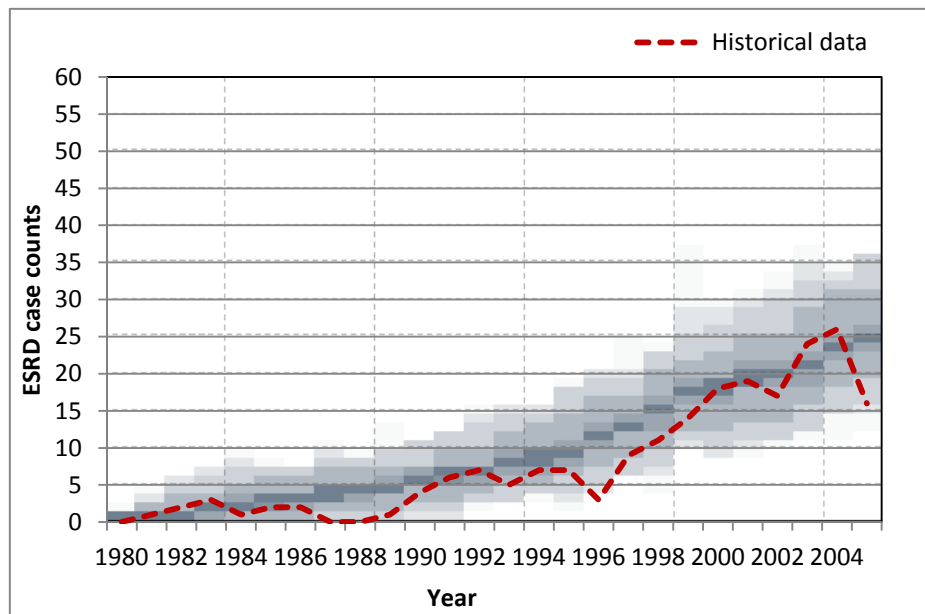


Figure 6.32 Comparison between incident historical data and model-estimated data for ESRD in OSK females from a Monte Carlo model.

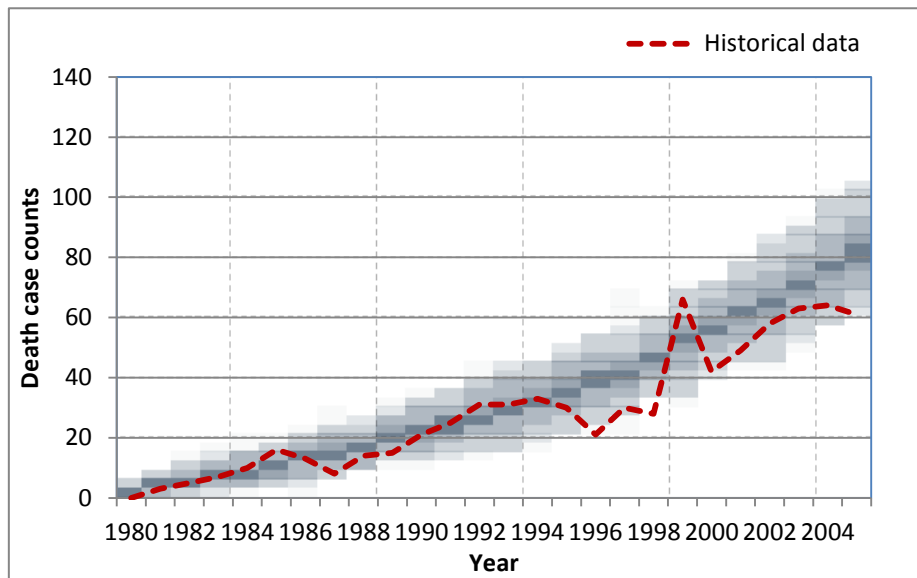


Figure 6.33 Comparison between incident historical data and model-estimated data for death without ESRD in FN males from a Monte Carlo model.

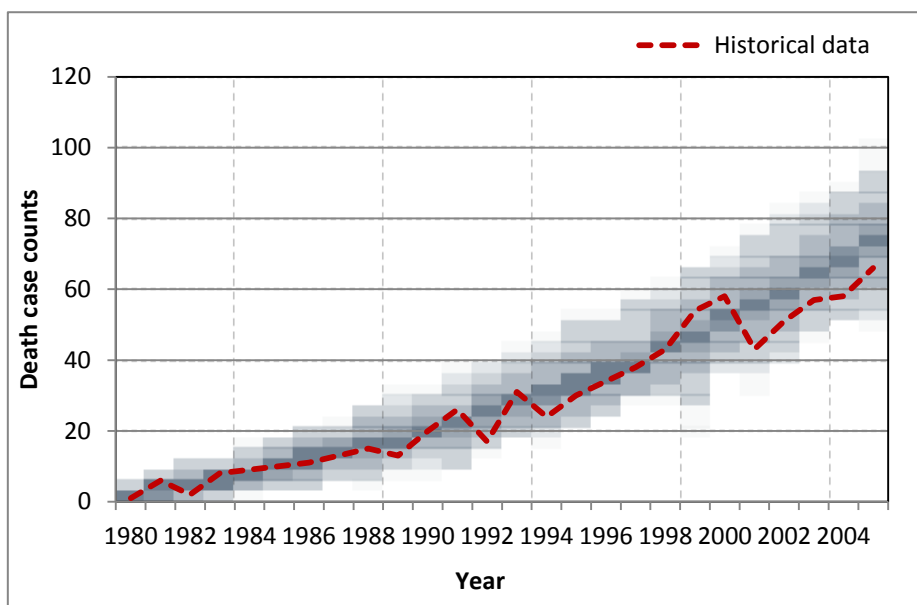


Figure 6.34 Comparison between incident historical data and model-estimated data for death without ESRD in FN females from a Monte Carlo model.

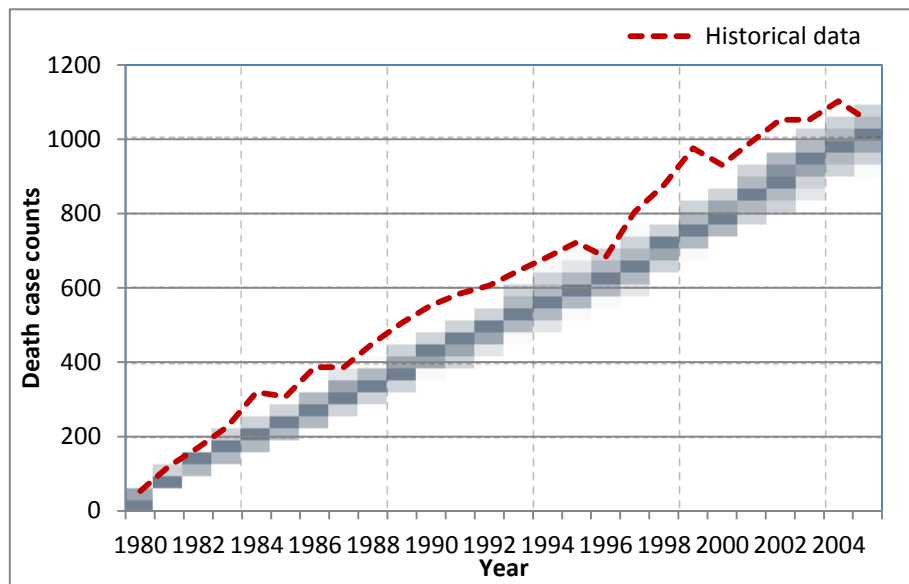


Figure 6.35 Comparison between incident historical data and model-estimated data for death without ESRD in OSK males from a Monte Carlo model.

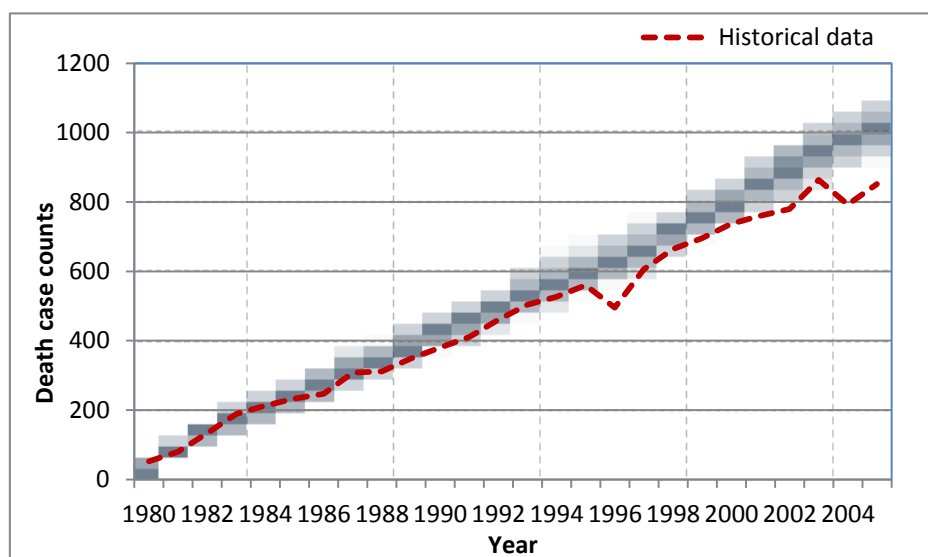


Figure 6.36 Comparison between incident historical data and model-estimated data for death without ESRD in OSK females from a Monte Carlo model.

## **Chapter 7 Discussion and conclusions**

### **7.1 Interpretation**

The results of this study will be discussed according to the study objectives:

1. To increase our understanding of why there are disparities in ESRD rates between FN and OSK with diabetes.
  - a. To determine whether there are significant disparities in the risk of ESRD development between FN and OSK.
  - b. To determine whether there are significant disparities in the risk of mortality without ESRD between FN and OSK.

First, the interpretation will start with the descriptive analysis. As observed in Figure 6.4, within the entire study cohort, the risk of death is higher than the risk of ESRD. FN people had higher hazard rates of ESRD compared with OSK over the whole study period (Figure 6.5). Figure 6.6 shows hazard rate curves for death without ESRD when not corrected for age of diabetes diagnosis, and among most of the study cohort, FN exhibited lower rates of death following diagnosis except after the 22nd year since diabetes diagnosis. To summarize, the risk of ESRD and the risk of death increase with time since diabetes diagnosis. FN had a higher risk of ESRD than OSK without adjusting for diabetes diagnosis age or sex; the crude risk of death without ESRD is lower among FN than that among OSK except near the end of study. Higher rates of death among OSK may play a role in the manifested lower rates of ESRD among OSK than FN.

Second, we will discuss the statistical modeling results. Cox cause-specific model analysis showed that FN people had about 3 times higher risk of ESRD than OSK people, after adjusting for diabetes diagnosis age and sex. Diabetes development at younger age increases the risk of ESRD. This may be because the long duration of diabetes leads to the accumulation of more damage to the kidneys (79). FN had lower risk of death than OSK before adjusting for age and sex difference. However, after adjusting for diabetes diagnosis age, sex, interaction between age and sex and interaction between age and ethnicity, FN had higher risk of death than OSK given the same sex and diabetes diagnosis age (younger than 81 years old). Moreover, most incident diabetes cases occur among FN in the 40-59 year age group, while most incident cases occur among OSK in the 70+ age group (23). From section 6.4, results from multivariable Cox cause-specific model for death without ESRD showed that after adjusting for 25-year age difference, FN had lower risk of death than OSK. Death as a competing risks event occurs more often among OSK, leading to the manifested situation that FN had higher rates of ESRD.

Similarly, parametric models can give hazard ratios similar to those of Cox cause-specific PH models. Moreover, exact estimations of hazard rates can be obtained based on different distributions associated with the hazard function, such as the Weibull distribution and piece-wise exponential distribution. From the Weibull PH model, we found that hazard rates for both ESRD and death increase over time since diabetes was diagnosed.

2. A lower risk of death without ESRD for FN than that for OSK has been shown in



Saskatchewan (7). The statistical analysis shows that with or without adjustment for age and sex, death risk is lower for FN than for OSK when patients are at the pre-ESRD stage. Moreover, similar competing risks of ESRD and death without ESRD have been shown in a comparison between African-Americans and non-African-Americans (51). When compared with non-African-Americans, African Americans were reported to have a higher ESRD risk for all age group and CKD stages and a higher risk of death with ESRD except for those being old and with severe kidney disease. Among Pima Indians, diabetes onset at a young age has been associated with increased rates of ESRD among Pima Indians. Longer accumulation of damage to the kidney may increase the risk of ESRD (79). One study reported the conflicting findings that FN had a lower crude risk of death compared with non-FN, while after adjustment for age and sex, FN had a higher risk of death (8).

3. To build Saskatchewan diabetic ESRD dynamic models.
  - c. To estimate hazard rates for ESRD and mortality without ESRD and to build a dynamic model using hazard rate estimations.
  - d. To validate models by comparing model-predicted data with historical data.

We used a piece-wise exponential model and built a System Dynamics model and an agent-based model based on piece-wise exponential distributions with two different intervals. To our understanding, the 3-year interval would be more precise because the baseline hazard rates of the 6th, 7th, and 8th intervals are different from the 9th interval. Here interval 6th indicates 15th year to 18th year since diabetes diagnosis.

As shown in Figure 6.31 to Figure 6.57, we found that there is a sizeable divergence between historical data and model-predicted data from the System Dynamics model but that the agent-based model matches the historic data quite closely, both in terms of point estimates and in terms of qualitative stochastic variability. The primary reason for this difference between SDM and ABM is that for the SDM, 3 years is the mean time for people to stay in the stock and then to transition to the next diabetes stock, and there might be some people who leave the stock quickly (without taking 3 years) and also some people who stay in the stock for more than 3 years, which distorts the estimation. However, in ABM, we track one individual flowing from one diabetes state to another, and that person stays in one diabetes state for precisely 3 years, which gives a good estimation of incident cases of ESRD and death without ESRD.

## **7.2 Limitations and strengths**

### **7.2.1 Limitations of the study population**

#### **Selection bias**

Selection bias is defined as “error due to systematic differences in characteristics between those who are selected for the study and those who are not” (80). One typical example of selection bias in a cohort study is loss to follow-up. Diabetes cases in this study were based on clinical diagnoses, which could have underestimated the frequency of diabetes. In addition, there might have been differences in screening and diagnosis strategies for people who have overt risk factors for diabetes, such as obesity, smoking, and high blood pressure, leading to differences in their inclusion in the

diabetic group.

### **Misclassification bias**

Misclassification bias, also called information bias, may “result from a systematic tendency for individuals selected for inclusion in the study to be erroneously placed in exposure or disease categories, thus leading to misclassification” (80). One type of misclassification bias comes from classification of races. We could only identify Aboriginal people as FN from the Indian Registry. Metis and non-registered Aboriginal people were included in the OSK group. This would tend to cause the current study to underestimate the real differences between FN and OSK in the risks of ESRD and death without ESRD.

### **Confounding and other risk factors**

The relationship among exposure variables, confounding factors and outcome variables is shown below (80, 81):

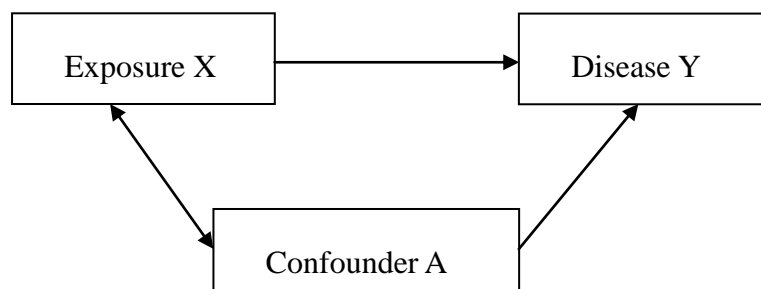


Figure 7.1 Confounding factor

Confounder A is an independent risk factor for disease, and it is not a mediator between X and Y (81). Moreover, confounder A is also related to exposure X but is not a result of exposure X (81). In our study, controlling for confounding factors was

limited. We controlled for age at diabetes diagnosis and sex in examining the differences in the risks of ESRD and death. This is a limited number of variables, which did not include several important risk factors for diabetes and ESRD. As mentioned in the literature review, risk factors for T2DM include obesity, physical inactivity, impaired glucose tolerance and GDM; for ESRD, risk factors include obesity, high blood pressure, GDM and poor glycemic control. Future research may need to explore other risk factors for both diabetes and ESRD and to compare the risks for ESRD and death without ESRD between groups.

### **7.2.2 Strengths of the study population.**

There are several advantages of the administrative dataset that we used, and some of them have been discussed in the literature (6, 23). First, the administrative database includes all of the covered population in Saskatchewan, which makes the study population a representative sample. Moreover, the sample size is large, which increases the study power. Surveys may underestimate the rates of diabetes based on their findings, suggesting that people tend to underreport diagnosed diabetes (80). Second, validated algorithms were used to define diabetes and ESRD. This increases the validity of the study. Third, the study period was 25 years, which is the longest, to our knowledge, in Canada. The long study period facilitates identifying cases of ESRD (6). All of the subjects in our study had diagnosed diabetes so that we could readily study diabetic ESRD among people who have diabetes.

### **7.3 Discussion of methodology.**

#### **Descriptive analysis**

For the descriptive analysis component of this thesis, we used both CIF and hazard rate graphs for both ESRD and death cases. Both CIF and cause-specific hazard rate curves give information regarding the study population (82). Cause-specific hazard curves help provide information on the underlying etiological mechanisms of events (83) and show the estimated instantaneous rate of an event per unit time. CIF can be understood as the probability of specific events happening before a certain time among the whole population (82).

#### **Modeling analysis**

In this research, we used semi-parametric and parametric models to for the survival data. The Cox cause-specific model can be modeled in SAS, and this method has been used to understand competing risk effects of death on ESRD among patients who have CKD (84). Derose et al. (51) used the Fine & Gray method (85) and found similar competing risks effects of death on ESRD rates. Lim et al. (83) compared these two methods and suggested that when the rates of competing risk events are different, it would be applicable to use a Cox cause-specific model instead of the Fine & Gray model. Moreover, Lim et al. (83) mentioned that when analyzing competing risk events, it is still useful to present both methods and compare the results.

Parametric models such as the Weibull PH model and piece-wise exponential PH model were tested to obtain exact estimates of hazard rates. The best-fitted Weibull

distribution model indicates that both hazard rates of ESRD and death increase over time. The Weibull PH model and piece-wise exponential PH model demonstrated that hazard ratios between groups are close to the hazard ratios from the Cox cause-specific model. The bridge between competing risks survival analysis and competing risks dynamic modeling is the hazard rate. Hazard rates for SDM for a given phase of disease progression (as represented by a stock) must be constant throughout the time since they entered the stock so that the chance for one person to leave is independent of how long they have been in that stock. Therefore, we decided to use a piece-wise exponential distribution model to derive the appropriate hazard rates. The time interval division was arbitrary, and further research is required to identify the optimal time interval division.

For the dynamic modeling investigation, we also compared ABM and SDM. Several comparisons can be discussed for these two methods of modeling (15, 66). First, for a large population and modest heterogeneity, ABM takes more time and computational memory than SDM. For example, 100 Monte Carlo simulations of the agent-based model took us almost half an hour, while an aggregate-model run took less than one minute. Moreover, more computer memory was required for ABM than that for the aggregate model. Second, if we double the size of the diabetes population, the increase of computational resources would increase at least linearly (66) in the agent-based model; by contrast, the performance of the System Dynamics model would remain the same. Although the ABM method takes more time and computational resources, there are several advantages to this method compared with aggregate modeling. First, in the aggregate model we used, subgroups represented population

characteristics, such as FN, OSK, males and females. We also used subgroups to represent diabetes diagnosis age (from 20 to 107 years). However, in ABM, capturing such heterogeneity can be performed with minimal performance cost and requires only adding parameters, such as *REGIND*, *SEX* and *DIABAGE*, to each agent (rather than imposing subscribing changes across the entire model), along with any associated logic. Second, while this thesis did not use this flexibility, an agent-based model permits using continuous hazard functions, such as the Weibull hazard function, instead of the memory-less assumption imposed by stocks in SDM. Third, in ABM (which incorporates stochastics), we are able to compare the degree of variability from the simulation results with those from the empirical data.

Comparisons between model predictions and historical data (Figure 6.10 to Figure 6.13) show that there are biases for both ESRD and death without ESRD. There are several possible reasons for these biases: 1) estimations from piece-wise exponential models may not be accurate enough; 2) chance may be playing a role and lead to bias; 3) the inexact time in transferring between stocks in the aggregate model could lead to some bias in the real transferring time. For example, in the 3-year interval model, some people may transfer quickly among stocks, and it may take them less than 3 years to go to the next stock, while others take longer than 3 years to go to the next stock. This may cause distortions in the historical data. We used ABM to test the third hypothesis, and the results gave considerably better matches for both ESRD and death without ESRD. In ABM, if we set up the time for one patient to progress to next diabetes state as 3 years, it is not mean time. Instead, it indicates that after exact 3 years, if the patient has not

developed ESRD or died, that patient will progress to diabetes.

## **7.4 Conclusions**

First Nations people had a higher risk of ESRD before and after adjusting for age and sex. FN had lower risk of death than OSK before adjusting for age and sex difference. After adjusting for diabetes diagnosis age, sex, interaction between age and sex and interaction between age and ethnicity, FN had higher risk of death than OSK given the same sex and diabetes diagnosis age except for those people who got diagnosed of diabetes at the age of older than or equal to 81 years old. After adjusting for about 25-year difference in the age of diabetes diagnosis, FN had lower risk of death compared to OSK people. The much younger age of diabetes diagnosis among FN compared to OSK likely contributes to higher rates of ESRD because of a differential mortality effect since FN with diabetes are more likely to live long enough to develop ESRD. Moreover, using the same hazard rate estimations, ABM showed better matches between historical data and model-predicted data compared to SDM.

## **7.5 Future work**

There are several possible avenues for future research on this topic. First, we may use the Fine and Gray model (85) to test the risk difference in ESRD or death without ESRD between FN and OSK or between males and females. We could compare the Fine and Gray model with the Cox cause-specific model. Although there are several macros in SAS to estimate CIF curves and test the difference between groups of CIFs,



for this task it would be convenient to use R software because there is an R package named “*compsk*” available to perform several such analyses. Second, we can use our agent-based model for diabetic ESRD and perform more tests, such as treating the hazard rates of death for FN and OSK the same in the model and testing whether there is any difference in ESRD rates between these two groups even with identical rates of death without ESRD. Additionally, we could include other factors, such as obesity, GDM, physical activity, dialysis, and treatment cost effects. Thirdly, more can be done in the piece-wise exponential model, such as dividing time into several intervals other than 3 years and 5 years. For example, we may divide the 25 years into six 3-year intervals and four 2-year intervals. Another possible future direction is to change the agent-based model using essentially continuous hazard functions, such as the Weibull hazard function.

## Reference

1. Himmelfarb J, Sayegh MH. Chronic kidney disease, dialysis, and transplantation : companion to Brenner & Rector's the kidney. Philadelphia: Saunders/Elsevier; 2010.
2. Canadian Institute for Health Information. Canadian Organ Replacement Register Annual Report: Treatment of End-Stage Organ Failure in Canada, 2000 to 2009. 2011 (Ottawa, Ont.: CIHI, 2011).
3. Canadian Institute for Health Information. Canadian Organ Replacement Register Annual Report: Treatment of End-Stage Organ Failure in Canada, 1995 to 2004. 2006 (Ottawa, Ont.: CIHI, 2011).
4. Young T, Kaufert JM, McKenzie JK, Hawkins A, O'Neil J. Excessive burden of end-stage renal disease among Canadian Indians: a national survey. *American Journal of Public Health*. 1989;79(6):756.
5. Wilson R, Krefting L, Sutcliffe P, VanBussel L. Incidence and prevalence of end-stage renal disease among Ontario's James Bay Cree. *Canadian journal of public health / Revue canadienne de santé publique*. 1992;83(2):143.
6. Dyck R, Osgood N, Lin T, Gao A, Stang M. End Stage Renal Disease Among People with Diabetes: A Comparison of First Nations People and Other Saskatchewan Residents from 1981 to 2005 *Canadian Journal of Diabetes*. 2010;34(4):324-33.
7. Dyck RF, Sidhu N, Klomp H, Cascagnette PJ, Teare GF. Differences in glycemic control and survival predict higher ESRD rates in diabetic first nations adults. *Clinical and investigative medicine Médecine clinique et expérimentale*. 2010;33(6):E390.
8. Gao S, Manns BJ, Culleton BF, Tonelli M, Quan H, Crowshoe L, et al. Prevalence of chronic kidney disease and survival among aboriginal people. *Journal of the American Society of Nephrology*. 2007;18(11):2953.
9. Pohar SL, Johnson JA. Health care utilization and costs in Saskatchewan's registered Indian population with diabetes. *BMC Health Services Research*. 2007;7(1):126.
10. Osgood N. Health Modeling Using AnyLogic. 2011. Available:

<http://www.cs.usask.ca/faculty/ndo885/Classes/CMPT858Spring2011/index.html>

(accessed 2011 April).

11. Homer JB, Hirsch GB. System dynamics modeling for public health: background and opportunities. *Journal Information*. 2006;96(3).
12. Motohashi Y, Nishi S. Prediction of end-stage renal disease patient population in Japan by system dynamics model. *International journal of epidemiology*. 1991;20(4):1032-6.
13. Jones AP, Homer JB, Murphy DL, Essien JDK, Milstein B, Seville DA. Understanding diabetes population dynamics through simulation modeling and experimentation. *Journal Information*. 2006;96(3).
14. Milstein B, Jones A, Homer JB, Murphy D, Essien J, Seville D. Charting Plausible Futures for Diabetes Prevalence in the United States: A Role for System Dynamics Simulation Modeling. *Preventing Chronic Disease*. 2007;4(3).
15. Nathaniel Osgood (2009). Topics in Modeling and Operations Research. [PowerPoint slides].
16. Young TK. The health of Native Americans: toward a biocultural epidemiology: Oxford University Press, USA; 1994. p. 5-8.
17. WHO. Definition, diagnosis and classification of diabetes mellitus and its complications. *Diagnosis and classification of diabetes mellitus*.99(2).
18. World Health Organization. Geneva (Switzerland): The Organization: World Health Organization; 2011. Available: <http://www.who.int/diabetes/en/> (accessed 2011 Feb.10).
19. International Diabetes Federation. The Diabetes Atlas. Brussels: The Federation: International Diabetes Federation; 2011. Available: <http://www.diabetesatlas.org/content/acknowledgements> (accessed 2011 Feb.10).
20. Zhang P, Zhang X, Brown J, Vistisen D, Sicree R, Shaw J, et al. Global healthcare expenditure on diabetes for 2010 and 2030. *Diabetes research and clinical practice*. 2010;87(3):293-301.

21. Public Health Agency of Canada. Report from the National Diabetes Surveillance System: Diabetes in Canada. Ottawa (ON): The Agency: Public Health Agency of Canada;2011.Available:<http://www.phac-aspc.gc.ca/publicat/2009/ndssdic-snsddac-09/index-eng.php> (accessed 2011 Feb.10).
22. First Nations Centre. First Nations Regional Longitudinal Health Survey (RHS) 2002/2003. Ottawa (ON): The Center: First Nations Centre; 2005. Available: <http://www.rhs-ers.ca/english/> (accessed 2011 Feb.10).
23. Dyck R, Osgood N, Lin TH, Gao A, Stang MR. Epidemiology of diabetes mellitus among First Nations and non-First Nations adults. Canadian Medical Association Journal. 2010;cmaj. 090846v1.
24. Hemmelgarn BR, Toth EL, King M, Crowshoe L, Ralph-Campbell K. Diabetes and First Nations people in Alberta. Alberta Diabetes Atlas 2007.127-40.
25. Green C, Blanchard JF, Young TK, Griffith J. The epidemiology of diabetes in the Manitoba-registered First Nation population. Diabetes care. 2003;26(7):1993.
26. Holt RIG. Textbook of diabetes. Chichester, West Sussex: Wiley-Blackwell; 2010.
27. Harrison LC, Honeyman MC. Cow's milk and type 1 diabetes: the real debate is about mucosal immune function. Diabetes. 1999;48(8):1501.
28. Alberti K, Zimmet P, Shaw J. International Diabetes Federation: a consensus on type 2 diabetes prevention. Diabetic Medicine. 2007;24(5):451-63.
29. Zimmet P. Type 2 (non-insulin-dependent) diabetes—an epidemiological overview. Diabetologia. 1982;22(6):399-411.
30. Zimmet P, Alberti K, Shaw J. Global and societal implications of the diabetes epidemic. Nature. 2001;414(6865):782-7.
31. Yusuf S, Hawken S, Ôunpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. The Lancet. 2004;364(9438):937-52.
32. Lee JM, Gebremariam A, Vijan S, Gurney JG. Excess Body Mass Index-Years, a Measure of Degree and Duration of Excess Weight, and Risk for Incident Diabetes.

Archives of Pediatrics and Adolescent Medicine. Arch Pediatr Adolesc Med. 2012;166(1):42-48.

33. Lau DCW, Douketis JD, Morrison KM, Hramiak IM, Sharma AM, Ur E. 2006 Canadian clinical practice guidelines on the management and prevention of obesity in adults and children [summary]. Canadian Medical Association Journal. 2007;176(8):S1.

34. Helmrich SP, Ragland DR, Leung RW, Paffenbarger Jr RS. Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. New England Journal of Medicine. 1991;325(3):147-52.

35. Schulz LO, Bennett PH, Ravussin E, Kidd JR, Kidd KK, Esparza J, et al. Effects of traditional and western environments on prevalence of type 2 diabetes in Pima Indians in Mexico and the US. Diabetes care. 2006;29(8):1866.

36. Dyck R, Klomp H, Tan LK, Turnell RW, Boctor MA. A comparison of rates, risk factors, and outcomes of gestational diabetes between aboriginal and non-aboriginal women in the Saskatoon health district. Diabetes care. 2002;25(3):487.

37. Berkowitz GS, Lapinski RH, Wein R, Lee D. Race/ethnicity and other risk factors for gestational diabetes. American Journal of Epidemiology. 1992;135(9):965.

38. Kim C, Berger DK, Chamany S. Recurrence of gestational diabetes: a systematic review. Diabetes care. 2007.

39. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes. Diabetes care. 2002;25(10):1862.

40. Creasy RK, Resnik R, Iams JD. Maternal-fetal medicine: principles and practice: WB Saunders Co; 2004.

41. Pettitt D, Nelson R, Saad M, Bennett P, Knowler W. Diabetes and obesity in the offspring of Pima Indian women with diabetes during pregnancy. Diabetes care. 1993;16(1):310.

42. Osgood ND, Dyck RF, Grassmann WK. The Inter-and Intragenerational Impact of Gestational Diabetes on the Epidemic of Type 2 Diabetes. American Journal of Public Health. 2011;101(1):173.

43. Tortora GJ, Derrickson BH. Principles of Anatomy and Physiology, 13th Edition Binder Ready Version: John Wiley & Sons; 2010.
44. Lerma EV. Evidence-Based Nephrology. JAMA: the journal of the American Medical Association. 2010;303(17):1758.
45. Zhang QL, Rothenbacher D. Prevalence of chronic kidney disease in population-based studies: systematic review. BMC Public Health. 2008;8(1):117.
46. CIHI. (2003). Health care in Canada. Ottawa: Canadian Institute for Health Information.
47. Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, Eggers P, et al. Prevalence of chronic kidney disease in the United States. JAMA: the journal of the American Medical Association. 2007;298(17):2038.
48. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu C. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. New England Journal of Medicine. 2004;351(13):1296-305.
49. Collins AJ, Li S, Gilbertson DT, Liu J, Chen SC, Herzog CA. Chronic kidney disease and cardiovascular disease in the Medicare population. Kidney international. 2003;64:S24-S31.
50. U.S. Renal Data System, USRDS 2007 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2007.
51. Derose SF, Rutkowski MP, Levin NW, Liu ILA, Shi JM, Jacobsen SJ, et al. Incidence of end-stage renal disease and death among insured African Americans with chronic kidney disease. Kidney international. 2009;76(6):629-37.
52. Dyck R, Tan L. Rates and outcomes of diabetic end-stage renal disease among registered native people in Saskatchewan. CMAJ: Canadian Medical Association Journal. 1994;150(2):203.

53. Nelson R, Newman J, Knowler W, Sievers M, Kunzelman C, Pettitt D, et al. Incidence of end-stage renal disease in type 2 (non-insulin-dependent) diabetes mellitus in Pima Indians. *Diabetologia*. 1988;31(10):730-6.
54. McClellan W, Warnock DG, McClure L, Campbell RC, Newsome BB, Howard V, et al. Racial differences in the prevalence of chronic kidney disease among participants in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) Cohort Study. *Journal of the American Society of Nephrology*. 2006;17(6):1710.
55. Nelson RG. Diabetic renal disease in transitional and disadvantaged populations. *Nephrology*. 2001;6(1):9-17.
56. Cass A, Cunningham J, Snelling P, Wang Z, Hoy W. Exploring the pathways leading from disadvantage to end-stage renal disease for Indigenous Australians. *Social Science & Medicine*. 2004;58(4):767-85.
57. Yamagata K, Ishida K, Sairenchi T, Takahashi H, Ohba S, Shiigai T, et al. Risk factors for chronic kidney disease in a community-based population: a 10-year follow-up study. *Kidney international*. 2006;71(2):159-66.
58. Pettitt D, Saad M, Bennett P, Nelson R, Knowler W. Familial predisposition to renal disease in two generations of Pima Indians with type 2 (non-insulin-dependent) diabetes mellitus. *Diabetologia*. 1990;33(7):438-43.
59. Dooley S, Metzger B, Cho N. Gestational diabetes mellitus. Influence of race on disease prevalence and perinatal outcome in a US population. *Diabetes*. 1991;40:25.
60. Rossing P, Hougaard P, Borch-Johnsen K, Parving HH. Predictors of mortality in insulin dependent diabetes: 10 year observational follow up study. *British Medical Journal*. 1996;313(7060):779.
61. Nelson RG, Morgenstern H, Bennett PH. Intrauterine diabetes exposure and the risk of renal disease in diabetic Pima Indians. *Diabetes*. 1998;47(9):1489.
62. Dyck RF. Tracking ancient pathways to a modern epidemic: diabetic end-stage renal disease in Saskatchewan aboriginal people. *Kidney international*. 2005;68:S53-S7.

63. Bonabeau E. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;99(Suppl 3):7280.
64. Hammond RA. Complex Systems Modeling for Obesity Research. *Preventing chronic disease*. 2009;6(3).
65. Kaplan G, Galea S, Riddle M. Complexity, epidemiology and the understanding of “what if”. *Journal of Epidemiology and Community Health*. 2011;65(Suppl 1):A41-A.
66. Tian Y, Osgood N. Comparison between Individual-based and Aggregate Models in the context of Tuberculosis Transmission. Accepted for Publication as a full paper in *Proceedings. The 29th International conference of the System Dynamics Society*. 2011;July 2011. Washington, D.C.
67. Hanley AJG, Harris SB, Mamakeesick M, Goodwin K, Fiddler E, Hegele RA, et al. Complications of type 2 diabetes among Aboriginal Canadians. *Diabetes care*. 2005;28(8):2054.
68. Tonelli M, Wiebe N, Culleton B, House A, Rabbat C, Fok M, et al. Chronic kidney disease and mortality risk: a systematic review. *Journal of the American Society of Nephrology*. 2006.
69. Kleinbaum DG, Klein M. *Survival analysis: a self-learning text*: Springer Verlag; 2005.
70. Berkson J, Gage RP. Calculation of survival rates for cancer. *Proceedings of the Staff Meetings of the Mayo Clinic*, 1950;25:270-86.
71. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*. 1958;53(282):457-81.
72. Collett D. *Modelling survival data in medical research*: CRC press; 2003.
73. SAS. <http://support.sas.com> (accessed Nov2011).
74. Breslow N. Covariance analysis of censored survival data. *Biometrics*. 1974;30(1):89-99.
75. Putter H, Fiocco M, Geskus R. Tutorial in biostatistics: competing risks and multi state models. *Statistics in Medicine*. 2007;26(11):2389-430.



76. Sterman JD. Business dynamics: Systems thinking and modeling for a complex world with CD-ROM: Irwin/McGraw-Hill; 2000.
77. Vensim. <http://www.vensim.com/subscript.html>. (accessed Nov 2010.).
78. Anylogic. [http://www.xjtek.com/anylogic/why\\_anylogic/](http://www.xjtek.com/anylogic/why_anylogic/). (accessed July 2011.).
79. Pavkov ME, Bennett PH, Knowler WC, Krakoff J, Sievers ML, Nelson RG. Effect of youth-onset type 2 diabetes mellitus on incidence of end-stage renal disease and mortality in young and middle-aged Pima Indians. JAMA: the journal of the American Medical Association. 2006;296(4):421.
80. Bonnie Janzen (2008). Epidemiology. [PowerPoint slides].
81. Gordis L. Epidemiology. Philadelphia: Elsevier/Saunders; 2009.
82. Pintilie M. Competing risks: a practical perspective: Wiley; 2006.
83. Lim H, Zhang X, Dyck R, Osgood N. Methods of competing risks analysis of end-stage renal disease and mortality among people with diabetes. BMC medical research methodology. 2010;10(1):97.
84. Agarwal R, Bunaye Z, Bekele DM, Light RP. Competing risk factor analysis of end-stage renal disease and mortality in chronic kidney disease. American journal of nephrology. 2008;28(4):569-75.
85. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American statistical association. 1999:496-509.